PCT/NL 2004 / 000229

Europäisches Patentamt | European Patent Office | Office européen des brevets

21. 06. 04

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03077307.1

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 03077307.1
Demande no:

Anmeldetag:
Date of filing: 18.07.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Vironovative B.V.
Burgemeester Oudlaan 50
3062 PA Rotterdam
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Sars-interferon

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

A61K38/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL PT RO SE SI SK TR LI

Title: SARS -interferon

The invention relates to the field of virology. More in particular the invention relates to the use of interferon, preferably pegylated interferon for the prophylactic or therapeutic treatment of animals, preferably vertebrates, more preferably birds or mammals, especially human, apes or rodents, infected with a coronavirus, more specifically an animal, preferably human infected with a SARS associated coronavirus (SARS-CoV).

Introduction

Recently, a new virus has caused a global health risk because of its pathogenic effects in man combined with a relatively easy droplet transmission. The virus first was seen in the Chinese province Guangdong, was spread to Hong Kong in February 2003, and within two months it has been able to spread to several countries all over the world where it has caused 78 deaths out of 2300 people infected (New Scientist Online News 13:25 02 April 2003). The virus has been named SARS (Severe Acute Respiratory Syndrome) virus and causes a respiratory illness (atypical pneumonia) in man. This illness usually begins with a fever, sometimes associated with chills or other symptoms, including headache, rash, diarrhea, a general feeling of discomfort (malaise) and body aches. Some people also experience mild respiratory syndromes at the outset.

After 2 to 7 days, SARS patients may develop a dry, nonproductive cough that might be accompanied or progress to the point where insuffiecient oxygen is getting to the blood, visible as shortness of breath. In 10% to 20% of the cases, patients will require mechanical ventilation, and eventually the disease can lead to the death of the patient. Hospital

personnel, children, elderly and people having an underlying condition such as diabetes or heart disease, or a weakened immune system, form the highest risk group. Co-infection with other pathogens seems to occur frequently, especially with opportunistic pathogenic microorganisms such as human metapneumovirus (hMPV), Chlamydia, etcetera.

The incubation time for the virus is typically 2-7 days and the disease is transmitted by people sick with SARS coughing or sneezing droplets in the air. SARS is a mammalian positive-sense single stranded RNA virus belonging to the Coronaviruses (e.g. isolate HK39849 EP patent application number 03076110.0, incorporated herein by reference). From a phylogenetic analysis of the sequences of the virus (Fig. 1) it appears that the virus is an intermediate between the group formed by TGEV (transmissable gastroenetritis virus), PEDV (porcine epidemic diarrhea virus) and 229E (human coronavirus 229E) at one side, the group formed by BoCo (bovine coronavirus) and MHV (murine hepatitis virus) at an other side, and the AIBV (avian infectious bronchitis virus) on yet another side. In general, bovine coronavirus seems to be the closest relative (at least for the viral replicase protein).

Coronaviruses were first isolated from chickens in 1937, while the first human coronavirus was propagated *in vitro* by Tyrell and Bonoe in 1965. There are now about 13 species in this family, which infect cattle, pigs, rodents, cats, dogs, birds and man. Coronavirus particles are irregularly shaped, about 60-220 nm in diameter, with an outer envelope bearing distinctive, 'club-shaped' peplomers (about 20 nm long and 10 nm wide at the distal end). This 'crown-like' appearance give the family its name. The envelope carries two glycoproteins: S, the spike glycoprotein which is involved in cell fusion and is a major antigen, and M, the membrane glycoprotein, which is involved in budding and envelope formation. The genome is associated with a basic phosphoprotein, designated N. The genome of coronaviruses, a single stranded positive-sense RNA strand, is

typically 27-31 Kb long and contains a 5' methylated cap and a 3' poly-A tail, by which it can directly function as an mRNA in the infected cell. Initially the 5' ORF 1 (about 20 Kb) is translated to produce a viral polymerase, which then produces a full length negative sense strand. This is used as a template to produce mRNA as a 'nested set' of transcripts, all with identical 5' non-translated leader sequence of 72 nucleotides and coincident 3' polyadenylated ends. Each mRNA thus produced is monocistronic, the genes at the 5' end being translated from the longest mRNA and so on. These unusual cytoplasmic structures are produced not by splicing, but by the polymerase during transcription. Between each of the genes there is a repeated intergenic sequence – AACUAAAC – which interacts with the transcriptase plus cellular factors to splice the leader sequence onto the start of each ORF. In some coronaviruses there are about 8 ORFs, coding for the proteins mentioned above, but also for a heamagglutenin esterase (HE), and several other non-structural proteins.

Newly isolated viruses are phylogenetically corresponding to and thus taxonomically corresponding to a SARS virus when comprising a gene order and/or amino acid sequence and/or nucleotide sequence sufficiently similar to our prototypic SARS virus. The highest amino acid sequence homology, between SARS virus and any of the known other viruses of the same family to date (BoCo or Mouse Hapatitis Virus) is for parts of the polymerase protein 18-61% (the % homology, and the virus to which the homology is depend on the region of the polymerase that is examined), as can be deduced when comparing the sequences given in figure 2 with sequences of other viruses, in particular of BoCo and Mouse Hapatitis Virus. Individual proteins or whole virus isolates with, respectively, higher homology than these mentioned maximum values are considered phylogenetically corresponding and thus taxonomically corresponding to SARS virus, and generally will be encoded by a nucleic acid sequence structurally corresponding with a sequence as shown in figure 2, and/or in GenBank accession no. AY274119 or AY278741 or AY338175 or AY338174

or AY322199 or AY 322198 or AY322197 or AH013000 or AY322208 or AY322207 AY 322206 or AY322205 or AH012999 and and/or sequences depicted in http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&id=227859&lvl=3&keep=1&srchmode=1&unlock, incorporated herein by reference. Herewith the invention encompasses a virus phylogenetically corresponding to the isolated virus of which the sequences are depicted in figure 2 and/or for example the GenBank accession no. AY274119 or AY278741 or AY338175 or AY338174 or AY322199 or AY 322198 or AY322197 or AH013000 or AY322208 or AY322207 AY 322206 or AY322205 or AH012999 and and/or sequences depicted in http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&id=227859&lvl=3&keep=1&srchmode=1&unlock, incorporated herein by reference.

"Interferon" is a term generically comprehending a group of vertebrate glycoproteins and proteins which are known to have various biological activities, such as antiviral, antiproliferative, and immunomodulatory activity at least in the species of animal from which such substances are derived. Interferon refers to a class of small protein and glycoprotein cytokines (15–28 kD) produced by T cells, fibroblasts, and other cells in response to viral infection and other biological and synthetic stimuli. Interferons bind to specific receptors on cell membranes; their effects include inducing enzymes, suppressing cell proliferation, inhibiting viral proliferation, enhancing the phagocytic activity of macrophages, and augmenting the cytotoxic activity of T lymphocytes. Interferons are divided into five major classes (alpha, beta, gamma, tau, and omega) and several subclasses (indicated by Arabic numerals and letters) on the basis of physicochemical properties, cells of origin, mode of induction, and antibody reactions.

Interferon has been the subject of intensive research on a worldwide basis. The literature is replete with publications concerning the synthesis of interferon, its proposed molecular characterizations, its clinical applications and proposed mechanisms of its antitumor, antiviral, and immune system activities. Because of the intensity and disparate origins of research concerning interferon and its characteristics and uses, there exists a substantial lack of uniformity in such matters as classification of interferon types. There are also numerous, sometimes contradictory, theories concerning the mode of action of interferon in producing clinical effects. Although originally isolated from cells of avian origin (chick allantoic cells), interferon production has been observed in cells of all classes of vertebrates, including mammals, amphibians, birds and reptiles. Interferon production by vertebrate cells is seldom spontaneous but is often readily "induced" by treatment of cells (in vivo or in vitro) with a variety of substances including viruses, nucleic acids (including those of viral origin as well as synthetic polynucleotides), lipopolysaccharides, and various antigens and mitogens.

Interferons have generally been named in terms of the species of animal cells producing the substance (e.g., human, murine, or bovine), the type of cell involved (e.g., leukocyte, lymphoblastoid, fibroblast) and, occasionally, the type of inducing material responsible for interferon production (e.g., virus, immune). Interferon has been loosely classified by some researchers according to induction mode as either Type I or Type II, with the former classification comprehending viral and nucleic acid induced interferon and the latter class including the material produced as a lymphokine through induction by antigens and mitogens. More recently, the international committee devising an orderly nomenclature system for interferon has classified interferon into types on the basis of antigenic specificities. In this newer classification, the designations alpha (.alpha.), beta (.beta.), and gamma (.gamma.) have been used to correspond to previous designations of leukocyte, fibroblast, and type II (immune) interferons,

respectively. Alpha and beta interferons are usually acid-stable and correspond to what have been called type I interferons; gamma interferons are usually acid-stable and correspond to what has been called type II interferons.

Alpha interferons have found the widest application in medicine. Alpha interferons are used in the treatment of chronic hepatitis B and hepatitis C, hairy cell leukemia, chronic myelogenous leukemia, AIDS-related Kaposi sarcoma, malignant melanoma, condylomata acuminata and recurrent respiratory papillomatosis due to human papillomavirus, and infantile hemangiomatosis. About 50% of patients treated for chronic hepatitis B with interferon-alfa show disappearance of hepatitis B_e antigen (HB_eAg) and reversion of alanine aminotransferase to normal. The response rate in chronic hepatitis C is lower (15–25%), but better results are achieved by using more aggressive therapy (daily rather than thrice weekly administration) and continuing it longer (a minimum of 12 months). Beta interferons reduce clinical recurrences and progression of myelin damage in multiple sclerosis. Gamma interferon is effective in retarding tissue changes in osteopetrosis and systemic scleroderma and in reducing the frequency and severity of infections in chronic granulomatous disease. Administration of interferons is parenteral (intravenous, intramuscular, subcutaneous, intranasal, intrathecal, or intralesional) and several weeks of treatment may be required before clinical response is noted. More than 50% of patients experience a flu-like syndrome of fatigue, myalgia, and arthralgia. Gastrointestinal and CNS side effects are also common, and marrow suppression may occur with prolonged treatment.

Modified formulations of interferon-alfa conjugated with polyethylene glycol (PEG) exist. Pegylated interferon alfa-2b (Peg-Intron) and alfa-2a (Pegasys) vary from the other interferons by having molecules of polyethylene glycol (PEG) attached to them. Pegylated IFN alfa-2b is

formed by covalent conjugation of a 12-kd monomethoxy polyethylene glycol (PEG) molecule to IFN alfa-2b, and pegylated IFN alfa-2a by covalent conjugation of a 40-kd branched mono-methoxy PEG molecule to IFN alfa-2a. Pegylation of the interferon molecule increases its size; the absorption of the larger pegylated molecule is slower, its half-life is longer, and its rate of clearance from the plasma is lower than that of the native interferon. Thus, the pegylated molecule increases the duration of biologic activity. The biological effect of "pegylation" is a prolonged presence in the serum (serum half life) and greater overall effect of the drug because of prolonged higher serum concentrations (increased AUC, area under the curve of concentration over time).

To date there is no effective treatment available for coronavirus diseases. The outbreak of SARS, which is a novel coronavirus and a highly infectious agent, warrants the search for antiviral compounds to treat SARS and other coronavirus related diseases, or newly emerging coronavirus infectious diseases. Coronaviruses cause economically important diseases of livestock, poultry, and laboratory rodents. Most coronaviruses of animals infect epithelial cells in the respiratory and/or enteric tracts, causing epizootics of respiratory diseases and/or gastroenteritis with short incubation periods (2-7 days), such as those found in SARS. In general, each coronavirus causes disease in only one animal species. In immunocompetent hosts, infection elicits neutralizing antibodies and cell-mediated immune response that kill infected cells. In SARS patients, neutralizing antibodies are detected 2-3 weeks after the onset of disease, and 90% of patients recover without hospitalization. In animals, reinfection with coronaviruses is common, with or without disease symptoms.

Until the present invention no specific treatment had been identified for coronavirus diseases of animals, livestock, and poultry and SARS-associated coronavirus infections. Several antiviral therapies have been

applied, but with various results. The present invention discloses that interferon, more specifically pegylated interferon can be used as a consistently effective prophylactic and therapeutic agent for treatment of coronavirus diseases, more specifically a disease caused by a SARS-associated coronavirus.

Summary of the invention

The invention provides use of interferon for the preparation of a medicament for the treatment or prevention of a coronavirus associated disease. Preferably said interferon is interferon-alpha, either interferon-alpha 2a or interferon-alpha 2b. Even more preferred said interferon is pegylated. For instance said coronavirus associated disease is a disease of animals, preferably vertebrates, more preferably birds or mammals, especially human, apes or rodents. For example said coronavirus associated disease is a respiratory disease and/or gastroenteritis. Preferably said animal is human. In a preferred embodiment said coronavirus associated disease is a disease caused by the feline infectious peritonitis virus (FIPV) or hemagglutinating encephalomyelitis virus (HEV) of swine or avian infectious bronchitis virus (IBV) or mouse hepatitis virus (MHV).

In another preferred embodiment said coronavirus associated disease is a disease caused by a SARS coronavirus (SARS-CoV). For instance, said SARS virus is a positive-sense single stranded RNA virus (SARS coronavirus) comprising one or more of the sequences of figure 2. For example, said SARS virus belongs to the Coronaviruses and identifiable as phylogenetically corresponding thereto by determining a nucleic acid sequence of said virus and testing it in phylogenetic tree analyses wherein maximum likelihood trees are generated using 100 bootstraps and 3 jumbles and finding it to be more closely phylogenetically corresponding to a virus isolate having the sequences as depicted in figure 2 than it is

corresponding to a virus isolate of BoCo (bovine coronavirus), MHV (murine hepatitis virus), AIBV (avian infectious bronchitis virus), PEDV (porcine epidemic diarrhea virus), TGEV (transmissible gastroenteritis virus) or 229E (human coronavirus 229E).

On the other hand said SARS virus is a positive-sense single stranded RNA virus (SARS coronavirus) corresponding to GenBank accession no. AY274119 or AY278741 or AY338175 or AY338174 or AY322199 or AY 322198 or AY322197 or AH013000 or AY322208 or AY322207 or AY 322206 or AY322205 or AH012999.

In one aspect the invention provides a method for the treatment or prevention of a coronavirus associated disease in an animal, preferably a vertebrate, more preferably a bird or mammal, especially human, ape or rodent, infected with a coronavirus infected with a coronavirus, said method comprising administrating interferon to said animal, preferably a vertebrate, more preferably a bird or mammal, especially human, ape or rodent, along with a pharmaceutically acceptable carrier.

Preferably said interferon is interferon-alpha, either interferon-alpha 2a or interferon-alpha 2b. Even more preferred interferon is pegylated. In another aspect the invention provides a method wherein said interferon is administered together with a vaccine and/or antiviral agent. Preferably said anti-viral agent is selected from the group consisting of attenuated vaccines, sub-unit vaccines, recombinant vaccines, antibody vaccines, nucleoside analogs such as ribavirin). In yet other aspect said interferon is administered in a dosage form adapted to assure maximum contact of the interferon in said dosage form with the oral and pharyngeal mucosa of an animal, preferably a vertebrate, more preferably a bird or mammal, especially human, ape or rodent.

Detailed description of the invention

The invention now provides the use of interferon for the prevention and the treatment of coronavirus mediated diseases. Preferably the coronavirus is SARS virus, but treatment of other coronaviruses (such as bovine coronavirus, murine hepatitis virus, avian infectious bronchitis virus, porcine epidemic diarrhea virus, transmissible gastroenteritus virus or human coronavirus 229E) is also part of the invention.

Most preferably the SARS virus is the virus having the sequences as disclosed in figure 2, and/or in GenBank accession no. AY274119 or AY278741 or AY338175 or AY338174 or AY322199 or AY 322198 or AY322197 or AH013000 or AY322208 or AY322207 AY 322206 or AY322205 or AH012999 and and/or sequences depicted in http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&id=227859&lvl=3&keep=1&srchmode=1&unlock, incorporated herein by reference, or a virus phylogenetically corresponding thereto.

Phylogenetically related viruses can be determined by testing a virus in phylogenetic tree analysis wherein maximum likelihood trees are generated using 100 bootstraps and 3 jumbles and finding it to be more closely phylogenetically corresponding to a virus having the sequences as depicted in figure 2 and/or in GenBank accession no. AY274119 or AY278741 or AY338175 or AY338174 or AY322199 or AY 322198 or AY322197 or AH013000 or AY322208 or AY322207 AY 322206 or AY322205 or AH012999 and and/or sequences depicted in http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&id=227859&lvl=3&keep=1&srchmode=1&unlock, incorporated herein by reference, than it is corresponding to a virus isolate of bovine coronavirus, murine hepatitis virus, avian infectious bronchitis virus, porcine epidemic diarrhea virus, transmissible gastroenteritus virus or human coronavirus 229E.

Another method of defining SARS viruses as a group is to group all viruses which have a higher percentage of homology to the sequences depicted in figure 2 and/or in GenBank accession no. AY274119 or AY278741 or AY338175 or AY338174 or AY322199 or AY 322198 or AY322197 or AH013000 or AY322208 or AY322207 AY 322206 or AY322205 or AH012999 and and/or sequences depicted in http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undef&id=227859&lvl=3&keep=1&srchmode=1&unlock, incorporated herein by reference, than virus-isolates of bovine coronavirus, murine hepatitis virus, avian infectious bronchitis virus, porcine epidemic diarrhea virus, transmissible gastroenteritus virus or human coronavirus 229E would have.

In general all interferon forms would be useful in the present invention, since it is known that all the interferon forms have at least some activity in alleviating (symptoms of) viral infection. However, it is to be understood that preferentially the interferon is used which is derived from the host which is infected with, or which runs the risk of being infected with the virus. Further, most preferred is the use of interferon-alpha, and especially – for coronavirus infections that affect humans, like SARS – human interferon-alpha. Alpha interferon is a natural protein produced by the human body in response to infection. It is also known as interferon alpha-2b. The type I interferon alpha family consists of small proteins that have clinically important anti-infective and anti-tumor activity. It is understood that alpha interferon may be administered alone or in combination with beta interferon or gamma interferon.

Genetic engineering techniques have allowed several companies to mass-produce alpha interferon, which is known as recombinant human alpha interferon, or by abbreviations such as rhIFN or rIFN-alpha. This is marketed under tradenames such as *Viraferon* (made by Schering-

Plough), *Roferon-A* (by Roche) and *Wellferon* (by Glaxo SmithKline). Interferon-alpha N3, or *Alferon N*, is another form of interferon alpha, derived from human leukocytes and containing multiple species of interferon-alpha.

A drawback to the use of interferon-alpha as discussed previously, is the short serum half-life and rapid clearance of the interferon alpha protein. However, it has been shown that the attachment of molecules of polyethylene glycol (PEG) to the interferon, creates a barrier that shields the interferon alfa-2a molecule from being rapidly degraded by proteases in the body and maintains its ability to consistently suppress the target virus over a longer dosage period.

As already discussed above pegylation of proteins such as inteferon is used to prevent rapid removal from the bloodstream and eventually rapid breakdown of the drug. A prolongation of the serum half-life of more than a factor two has been demonstrated (Shannon A. Marshall, Drug Discovery Today Volume 8, Issue 5 , March 2003, Pages 212-221). Pegylated IFN alfa-2b has a prolonged serum half-life (40 hours) relative to standard IFN alfa-2b (7–9 hours). The greater size of pegylated IFN alfa-2a acts to reduce glomerular filtration, markedly prolonging its serum half-life (72–96 hours) compared with standard IFN alfa-2a (6–9 hours) (Bruce A. Luxon MD Clinical Therapeutics, Volume 24, Issue 9, September 2002, Pages 1363-1383).

Pegylation of proteins is a standard technique available to a person skilled in the art, and standard pegylated interferons are available commercially Roche (PEGASYS® (interferon alfa 2a) and Schering-Plough (PEG-Intron A) or in development like PEG-Alfacon, the PEGylated version of Infergen(R) (Interferon alfacon-1) a bio-engineered type I interferon alpha. Schering-Plough has developed a semi-synthetic form of Intron® A by attaching a 12-kDa mono-methoxy polyethylene glycol to the protein (PEG

Intron) which fulfils the requirements of a long-acting interferon alpha protein while providing significant clinical benefits. Pegylation decreases the specific activity of the interferon alpha-2b protein, whilst the potency of PEG Intron, independent of protein concentration, is comparable to the Intron® A standard at both the molecular and cellular level. PEG Intron has enhanced pharmacokinetic profile in both animal and human studies [see Yu-Sen Wang et al., 2002: Advanced Drug Delivery Reviews, Volume 54, Issue 4 , 17, Pages 547-570]. In PEGASYS, a 40 kilodalton branched, mobile PEG is covalently bound to the interferon alfa-2a molecule and provides a selectively protective barrier without significantly reducing binding site receptivity.

It is understood that pharmacokinetic behaviour of a pegylated molecule depends on the size of the PEG and the structure of the link between the PEG moiety and the protein (Shannon A. Marshall, Drug Discovery Today Volume 8, Issue 5 , March 2003, Pages 212-221). It is known that interferons with smaller PEGs are degraded quickly, requiring more frequent dosing. Thus interferons with larger PEGs are preferred.

Thus the present invention encompasses all types of pegylated interferons or future interferons with yet undisclosed molecule attachments which provide a selectively protective barrier, to shield the interferon from being degraded, without significantly reducing binding site receptivity. Also combinations of different interferons are encompassed in the invention

One of the most preferred embodiments of the present invention is the use of interferon as a prophylactic treatment for the prevention of coronavirus infection. Subjecting apes to a prophylactic or therapeutic treatment either before or during infection with the cornoavirus has a good and useful predictionary value for application of such a prophylaxis or therapy in human subjects.

As is shown in the experimental section administration of interferon before infestation with virus particles greatly delays infection and the effects after infection. It should be understood that the virus challenge given to the test animals is a high dose, which will not or hardly ever occur in 'natural' infections. It is understood that viral challenge under 'natural' circumstances would equate with a challenge of about 10-10^5 TCID 50 with a concentration which is much less than that used in the experiment of the invention. Further, the viral challenge in the experiment was for the largest part applied intra-tracheal, i.e. at the place where the virus exerts its main infective activity. Normally, a virus will be encountered in the air that is breathed and this air will firstly pass the nose and/or oral cavity, where it will have a large chance of being filtered out (and stopped) by the epithelium and mucosa of the mouth and/or the nose. Anyhow, the fact that even at the extremely high dose used in our experiments we have been able to show effect of interferon indicates that the effect will even be more pronounced at infective viral doses which are normally encountered. It is therefore believed that prophylactic administration will give a durable and strong protection against infection with coronaviruses.

This is especially important in relation to viruses which are highly infective and/or which have an airborne mode of transmission, such as, for instance, the SARS virus. A prophylactic treatment would be especially welcome for people who run a risk of being infected, such as, in the case of SARS virus, hospital personnel, children, elderly and people having an underlying condition such as diabetes or heart disease, or a weakened immune system.

It remains possible that SARS-CoV infection might be asymptomatic in some people, or cause nonrespiratory symptoms in others. There is insufficient evidence to exclude the possibility that asymptomatic, or

atypical, infected people can transmit the disease. Thus a prophylactic treatment for the prevention of coronavirus infection, like SARS-CoV is indeed essential.

However, our data also show that interferon is also applicable for therapy of coronaviruses, i.e. at the time when virus infection is already established. Our *in vivo* data show that pathologic effects are at least delayed upon administration of interferon.

Interferon of human and murine origins has been quantified in the art in terms of International Units ("IU"). As used herein, a "unit" of interferon (to be distinguished from "IU") shall mean the reciprocal of a dilution of interferon-containing material that, as determined by assay, inhibits one-half the number of plaques of a challenge virus, the challenge virus being the vesicular stomatitis virus ("VSV"). So quantified a "unit" of interferon is routinely found to be about one-tenth the quantity of interferon represented by one "IU. " Alternatively, interferon can be quamtitated in µg/kgof body weight.

Interferon is given in doses ranging from 1µg/kg to 3µg/kg. When the interferon is pegylated doses can be delivered less frequently. Treatment of a coronavirus disease in accordance with the present invention comprises administering pegylated interferon at a dosage of 0.01-6µg/kg per day in a dosage form adapted to promote contact of said dosage of interferon with the oral and pharyngeal mucosa of said animal. Preferably, the dosage of interferon is from 0.1-4µg/kg per day, more preferably 0.3-3µg/kg per day.

Interferon may be administered by any available means, including but not limited to, oral, intravenous, intramuscular, pulmonary and nasal routes,

and wherein said composition is present as a solution, a suspension or an aerosol spray, especially of fine particles.

It is critical that the pegylated interferon be administered in a dosage form adapted to assure maximum contact of the interferon in said dosage form with the oral and pharyngeal mucosa of the human or animal, undergoing treatment. Contact of interferon with the mucosa can be enhanced by maximizing residence time of the treatment solution in the oral or pharyngeal cavity. Thus, best results seem to be achieved in human patients when the patient is requested to hold said solution of interferon in the mouth for a period of time. Contact of interferon with the oral and pharyngeal mucosa and thereafter with the lymphatic system of the treated human or animal avian, rodent is unquestionably the most efficient method administering immunotherapeutic amounts of pegylated interferon.

For example interferon can be administered in either a liquid (solution) or solid dosage form. Thus interferon can be administered dissolved in a buffered aqueous solution typically containing a stabilizing amount (1-5% by weight) of blood serums. Exemplary of a buffered solution suitable as a carrier of interferon administered in accordance with this invention is phosphate buffered saline prepared by standard techniques.

It is also contemplated by the present invention to provide interferon in a solid dosage form such as a lozenge adapted to be dissolved upon contact with saliva in the mouth with or without the assistance of chewing. Such a unitary dosage form is formulated to release about 1 to about 1500 IU of interferon upon dissolution in the mouth for contact with the oral and pharyngeal mucosa. Thus a unitary dosage form of interferon in accordance with this invention can be prepared by art-recognized techniques for forming compressed tablets such as chewable vitamins. Similarly, interferon can be incorporated into starch-based gel

formulations to form a lozenge which will dissolve and release interferon for contact with the oral mucosa when held in the mouth. Solid unitary dosage forms of interferon for use in accordance with the present invention can be prepared utilizing art recognized dosage formulation techniques. The pH of such formulations can range from about 4 to about 8.5. Of course, in processing to such unitary dosage forms one should avoid heating a pre-dosage form formulation, after addition of interferon, above about 50°C. Exemplary of a solid dosage form for animal use is a molasses block containing effective amounts of interferon.

Alternatively the interferon can be formulated into flavoured or unflavoured solutions or syrups using a buffered aqueous solution of interferon as a base with added caloric or non-caloric sweeteners, flavour oils and pharmaceutically acceptable surfactant/dispersants.

Also contemplated are methods of gene therapy capable of causing expression of interferon in respiratory or gastric cells for prevention of a coronaviral infection.

Of course, the clinical use of any medicament of the present invention is a clinical decision to be made by the clinician and the exact course of such treatment is left to the clinician's sound discretion, with all such courses of treatment deemed within the bounds of the present invention.

Another preferred embodiment is administration of interferon together with another treatment which is directed to prevent or treat infection with coronaviruses. Such other treatment can for instance be vaccination which can be used either to prevent or to treat infection. Vaccine production for coronaviruses can be done by active vaccination, using attenuated viruses, and/or vaccines composed of viral subunits (so called subunit vaccines), or by passive vaccination with antibodies active against coronaviruses.

For example a peptide comprising part of the amino acid sequence of the spike protein as depicted in translation 2 with the sequence EMC7 and translation 1 of the RDG seq of figure 2, can be used for the preparation of a therapeutic or prophylactic peptide. Also included are peptides comprising part of the amino acid sequence of the spike protein of a positive-sense single stranded RNA virus (SARS coronavirus) corresponding to GenBank accession no. AY274119 or AY278741 or AY338175 or AY338174 or AY322199 or AY 322198 or AY322197 or AH013000 or AY322208 or AY322207 or AY 322206 or AY322205 or AH012999 or yet to be isolated SARS coronaviruses.
Also included are peptides comprising part of the amino acid sequence of the spike protein of the feline infectious peritonitis virus (FIPV) or hemagglutinating encephalomyelitis virus (HEV) of swine or avian infectious bronchitis virus (IBV) or mouse hepatitis virus (MHV).

Preferably a protein comprising the amino acid sequence of the spike protein of the feline infectious peritonitis virus (FIPV) or hemagglutinating encephalomyelitis virus (HEV) of swine or avian infectious bronchitis virus (IBV) or mouse hepatitis virus (MHV) is used for the preparation of a sub-unit vaccine. Even more preferred a protein comprising the amino acid sequence of the spike protein of a positive-sense single stranded RNA virus (SARS coronavirus) corresponding to GenBank accession no. AY274119 or AY278741 or AY338175 or AY338174 or AY322199 or AY 322198 or AY322197 or AH013000 or AY322208 or AY322207 or AY 322206 or AY322205 or AH012999 or yet to be isolated SARS coronaviruses is used for the preparation of a sub-unit vaccine. Even more preferred a protein comprising the amino acid sequence of the spike protein as depicted in translation 2 with the sequence EMC7 translation 1 of the RDG seqof figure 2, is used for the preparation of a sub-unit vaccine. Furthermore, the nucleocapsid of Cornoviruses, as depicted in the translation of EMC8, in figure 2, is known to be particularly useful for eliciting cell-mediated immunity against

Coronaviruses and can also be used for the preparation of a sub-unit vaccine.

Attenuation of the virus can be achieved by established methods developed for this purpose, including but not limited to the use of related viruses of other species, serial passages through laboratory animals or/and tissue/cell cultures, serial passages through cell cultures at temperatures below 37C (cold-adaption), site directed mutagenesis of molecular clones and exchange of genes or gene fragments between related viruses.

Antibodies against coronavirus proteins, like SARS virus proteins, especially against the spike protein of coronaviruses, like the SARS virus, for example preferably against the amino acid sequence as depicted in translation 2 of EMC7 and translation 1 of the RDG seq in figure 2, are also useful for prophylactic or therapeutic purposes, as passive vaccines. It is known from other coronaviruses that the spike protein is a very strong antigen and that antibodies against spike protein can be used in prophylactic and therapeutic vaccination.

Use of interferon together with administration of a vaccine will boost the effects of the vaccine. First of all, there is the combination of treatments that will add up to a better effect. However, co-administration of interferon with a vaccine will also enable the immune response to vaccination to have more effect. Normally the immune response is slow and it takes a few days to come to a high enough titer of antibodies to be able to effectively combat virus particles. When no interferon is co-administered the virus would have had the chance to multiply to enormous amounts, which cannot be overcome by the immune response. With interferon, however, the amounts of the virus will remain absent or low and any infective virus outburst (if any at all) can easily be handled by the immune system.

Treatment to prevent and/or treat infection with coronaviruses can also comprise combination treatments with other antiviral compounds, such as, for instance, nucleoside-based compounds such as ribavirin (e.g. Rebetol® (ribavirin, USP). These compounds act through interfering with the viral replication by presenting nucleosides which are built in during viral replication, but which either prevent formation of viral proteins or which do not yield functional proteins. Co-administration of interferon will even more slow down viral replication. Combinations of ribavirin and forms of interferon can help to reduce viral load.

Another disease condition responding to treatment in accordance with the present invention is neoplastic disease. Thus, the administration of interferon in accordance with the above description can, alone or in combination with other drugs or therapy, help effect remission of cancers such as malignant lymphoma, melanoma, mesothelioma, Burkitt lymphoma and nasopharyngeal carcinoma and other neoplastic diseases, especially those of known or suspected viral etiology and diseases such as Hodgkin's Disease and leukemia.

Other disease conditions responding to treatment in accordance with the present invention are infectious diseases of coronaviral origin avian, porcine, canine and feline species. Several other coronaviruses can cause fatal systemic diseases in animals, including feline infectious peritonitis virus (FIPV), hemagglutinating encephalomyelitis virus (HEV) of swine, and some strains of avian infectious bronchitis virus (IBV) and mouse hepatitis virus (MHV). These coronaviruses can replicate in liver, lung, kidney, gut, spleen, brain, spinal cord, retina, and other tissues. Immunopathology plays a role in tissue damage in MHV and FIPV, and cytokines are responsible for some signs of disease. Significantly, in cats with persistent, inapparent infection with feline enterotropic coronavirus, virulent virus mutants can arise and cause fatal infectious peritonitis, a systemic disease.

Figure legends

Fig. 1: Phylogenetic relationship for the nucleotide sequences of isolate HK39849 with its closest relatives genetically. Phylogenetic trees were generated by maximum likelihood analyses using 100 bootstraps and 3 jumbles. The scale representing the number of nucleotide changes is shown for each tree.

Fig. 2: Nucleotide sequences from 13 clones of parts of the SARS virus. Also included are the putative polypeptide sequences of polypeptides and alignments of the putative polypeptides with that of another member of the Coronoviridae family, where possible.

Fig. 3: Schematic map of the SARS virus genome, indicating the position of the nucleotide sequences of figure 2 relative to the genome and a putative indication of the open reading frames of the genome based on analogy with other coronaviruses. The gene structure for the region between the Spike and Nucleocapsid is uncertain. EMC1-EMC14 and RDG 1: sequences as provided in figure 2. CDC and BIN1-2: sequences were provided through personal communication from the CDC (Dr. W. Bellini, Centers for Disease Control & Prevention, National Centers for Infectious Diseases, 1600 Clifton Road, Atlanta GA 30333, USA) and BNI (Dr. C . Drosten and Prof. Dr. H. Schmitz, Bernard Nocht Institute, Bernard-Nocht Str. 74, D-20359 Hamburg, Germany), respectively.

Fig. 4: Amino acid comparison of the N-terminus of the S-protein of the SARS virus and closely related coronaviruses. HCV OC43 = human coronavirus isolate OC43; MHV A59 = murine hepatitis virus isolate A59, BCV = bovine corona virus.

Fig. 5: Negative contrast EM photograph of SARS virus obtained from concentrated supernatant of infected cell cultures.

Fig. 6: Infection with SARS-coronavirus causes pulmonary and renal lesions in cynomolgus macaques. Formalin-fixed, paraffin-embedded tissues were stained with haematoxylin and eosin and examined by light microscopy. There is diffuse alveolar damage of the lung (a), and the alveolar lumina (b) are flooded with highly proteinaceous exudate admixed with inflammatory cells and cellular debris. In the lumen of a bronchiole (c) and in the surrounding lung parenchyma are several multinucleated syncytial cells (arrowheads). The renal collecting tubules (d) contain similar multinucleated syncytial cells. Original magnifications: a x 12.5; b x 50; c x 100; d x 250.

Examples

Virus isolation and characterisation

Isolate HK39849 was isolated from a hospitalised SARS patient by throat swab and inoculated into a culture of Vero-E6 cells. A sample of the supernatant from these infected cells was provided by Dr. M. Peiris (Queeen Mary Hospital Faculty of Medicine, Hong Kong University, Honk Kong) was used to inoculate VERO-118 cells and cell culture supernatant from these cells was aliquoted and frozen after one passage

We isolated RNA from the virus-containing cell culture supernatant and subjected it to RNA arbitrarily primed PCR (RAP-PCR) essentially as described by Welsh & McClelland (NAR 18:7213; PNAS USA 90:10710, 1993). Virus in the culture supernatants was purified on continuous 20-60% sucrose gradients. The gradient fractions were inspected for virus-like particles by EM, and RNA was isolated from the fraction containing , in which the most nucleocapsids were observed. Equivalent amounts of RNA isolated from virus fractions were used for RAP-PCR, after which samples were run side by side on a 3% NuSieve agarose gel. Differentially displayed bands ranging in size from 200-1500 base pairs specific for the unidentified virus were subsequently purified from the gel, cloned in plasmid pCR2.1 (Invitrogen) and sequenced with vector-specific primers. When we used these sequences to search for homologies against sequences in the Genbank database using the BLAST software (www.ncbi.nlm.nih.gov/BLAST/) which yielded resemblance to virus sequences of the coronaviruses displayed in the phylogenetic tree of figure 1.

Eight of these fragments (EMC 1-6, 13 and 14) were located in the ORF coding for the viral polymerase (ORF 1ab), one (EMC-7) spanned the 3' end of ORF1ab and reached into the 5' end of spike protein region; EMC-10 overlapped the 3' end of EMC-7 and therefore also codes part of the S protein region and EMC 9 encodes a region downstream of EMC-10; by use of primers to sequences within EMC10 and EMC9 (see below), the

region between these two sequences was amplified by PCR and sequenced. The full contiguous region has been incoporated into EMC7 in firgure 2;a further sequence (RDG1 in figure 2) encodes the 3' end of the Spike protein. A further sequence (EMC8) spanned part of the Nucleocapsid coding sequence. The remaining three sequences (EMC9, 11 and 12) encode regions of as yet unknown function.

Phylogeny

BLAST searches using nucleotide sequences obtained from the unidentified virus isolate revealed homologies primarily with members of the Coronaviridae. As an indication for the relation between the newly identified virus isolate and other coronaviruses a phylogenetic tree was constructed based on the sequence information obtained (figure 1).

Phylogenetic analyses

For all phylogenetic trees, DNA sequences were alligned using the ClustalW software package and maximum likelihood trees were generated using the DNA-ML software package of the Phylip 3.5 program using 100 bootstraps and 3 jumbles[15]. Previously published sequences for TGEV, PEDV, 229E, AIBV, BoCo and MHV that were used for the generation of phylogenetic trees are available from Genbank

SARS- interferon experiment

Four groups of two monkeys were injected.

1. *PEG-INTERFERON treatment*

   Dose: 3 μg/kg or PBS injected intramuscularly according the following scheme:

   Monkey: M001 PBS at days –3, -1, +1 and +3

M002	PBS	at days –3, -1, +1 and +3
M003	IFN	at days –3, -1, +1 and +3
M004	IFN	at days –3, -1, +1 and +3
M005	PBS	at d. –3 and –1 and IFN at d. +1 and +3
M006	PBS	at d. –3 and –1 and IFN at d. +1 and +3
M007	IFN	at days –3, -1, +1 and +3
M008	IFN	at days –3, -1, +1 and +3

2. *Infection*

SARS coronavirus infection of all monkeys on day 0

Dose: 10^6 TCID 50 in 5 ml PBS

- 4 ml intra-tracheal
- 1 ml intranasal
- 0.5 ml on each of the eyes

3. *Sampling*

a. Nose throat and rectum swabs taken on days 0, 2 and 4 and were put in 1 ml transport medium.

b. Monkeys were euthanised on day 4 and samples of lung, tracheal bronchial lymph node and trachea were harvested

Virus was cultured and titrated on Vero-118 cells, and these were scored for cytopathic affects

Virus titration using the three different swabs taken on days 0 , 2 and 4 after infection (nose, throat and rectum) and isolation of virus from the lungs, tracheal bronchial lymph node and trachea at day 4 after infection demonstrated that the two control monkeys (M001 and M002) were successfully infected (table 1).

Table 1 SARS-associated coronavirus excretion by cynomolgus macaques treated with pegylated interferon.

Animal no.	specimen*					
	Pharyngeal swab			Tr. Br lymph node	Trachea	Lung
	0	2	4	4	4	4
M001	-	+	+	+	+	+
M002	-	+	+	+	+	++
M003	-	-	-	-	-	+
M004	-	-	-	+	-	+
M005	-	+	-	-	+	++
M006	-	+	-	+	+	++
M007	-	-	-	n.a.	n.a.	n.a.

M008	-	-	-	n.a.	n.a.	n.a.

* day post infection

1. *Control animals (M001, 002)*
   - Pharyngeal swabs on days 2 and 4 were all positive
   - Animal M001 also was found positive with respect to isolation of SARS coronavirus from the nasal swab (day 2 and 4).
   - No rectal swabs were positive
   - Tissue specimen from the lungs, trachea and trachea bronchial lymph node from both control animals (M001 and M002) were positive at day 4 when the animals were sacrificed. The lung tissue homogenate contained virus at a high titer because the Vero cultures were found positive rapidly after inoculation.

2. *Prophylactically treated animals (M003, 004, 007 & 008)*

   - negative with respect to the virus isolation test on pharyngeal swabs taken at day 0, 2 and 4 after infection (table 1).
   - No nasal swab was found positive in these animals.
   - Only one rectal swab of animal M004 at day 4 was scored positive (which has to be confirmed in the PCR assay because these cultures showed much bacterial contamination (cultures of rectal swabs)
   - No virus isolated from trachea of M003 and 004
   - Virus isolated from tracheal bronchial lymph node of M004 but not M003
   - Virus isolated from lungs of M003 and 004, but are at lower titre than controls as it took longer for CPE to be observed in Vero-118 cells inoculated with samples from the lungs (confirmed by PCR – figure A below)

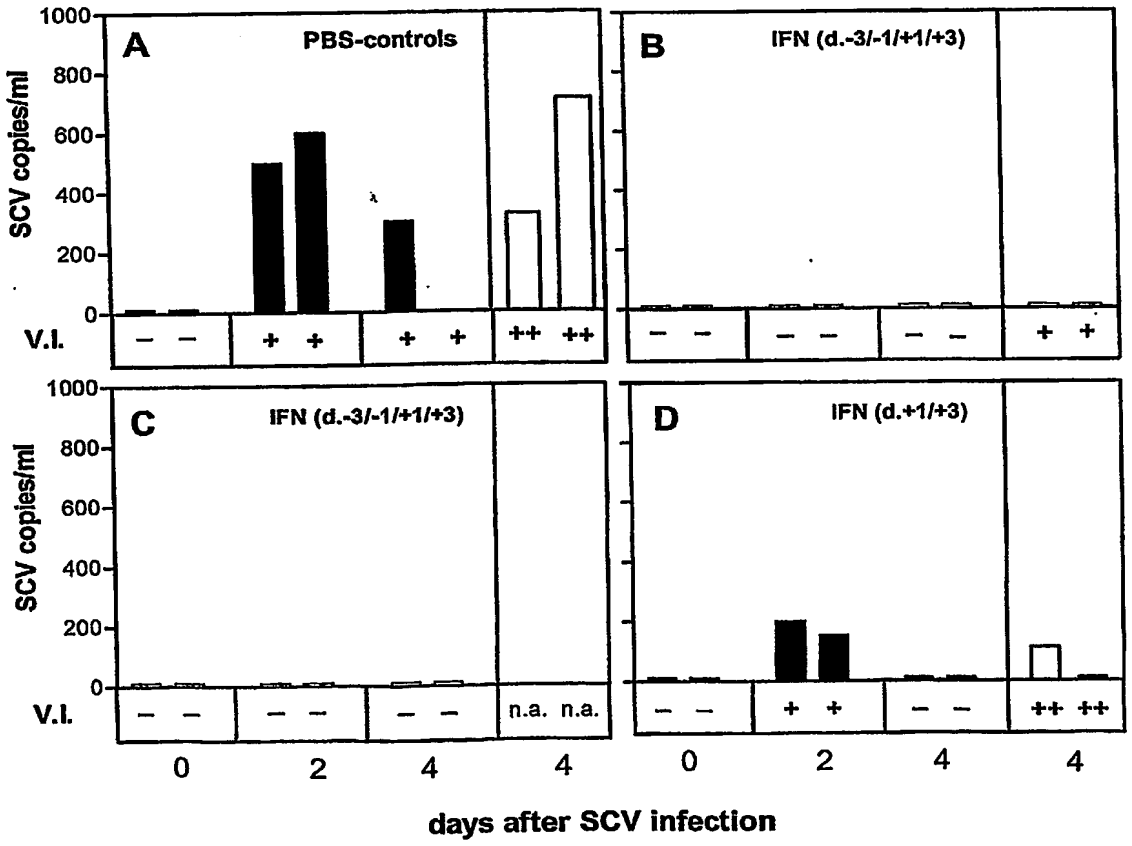
3. *Therapeutically treated animals (M005 and 006)*

SARS coronavirus

- isolated from pharyngeal swabs taken at day 2 after infection
- not isolated from the pharyngeal swabs taken at day 4 after infection.
- isolated from more tissue samples and at higher titers from animal M005 and M006, than from animal M003 and M004 (quantitation confirmed by PCR)

Pathological examination of lung section stained by HE confirmed the low level infection of the lungs of animal M003.

Figure A Effect of pegylated IFN-α on SARS Coronavirus (SCV) replication in macaques. SCV detection in pharyngeal swabs (days 0, 2 and 4 after infection, closed bars) and lungs (day 4, open bars) taken from



cynomolgus monkeys treated with PBS (A), PEG-Intron at days –3, -1, +1 and +3 (B and C) and PEG-Intron at days +1 and +3 (D) after SCV infection. Individual macaques are shown (n=2 per group). Virus isolation (VI) results are indicated in the lower part of the panel whereas real time PCR results are shown in the upper part of the panels (n.a., not available).

Claims

1. Use of interferon for the preparation of a medicament for the treatment or prevention of a coronavirus associated disease.

2. Use according to claim 1 wherein said interferon is interferon alpha.

3. Use according to claim 2, wherein said interferon is interferon-alpha 2a.

4. Use according to claim 2, wherein said interferon is interferon-alpha 2b.

5. Use according to any of claims 1-4, wherein said interferon is pegylated.

6. Use according to any of claims 1-5, wherein said coronavirus associated disease is a disease of animals, preferably vertebrates, more preferably birds or mammals, especially humans, ape or rodent.

7. Use according to claim 6, wherein said disease is a respiratory disease and/or gastroenteritis.

8. Use according to claim 6 or claim 7, wherein said animal is human.

9. Use according to any of claims 1-7 wherein said coronavirus associated disease is a disease caused by the feline infectious peritonitis virus (FIPV) or hemagglutinating encephalomyelitis

virus (HEV) of swine or avian infectious bronchitis virus (IBV) or mouse hepatitis virus (MHV).

10. Use according to any of claims 1-7 wherein said coronavirus associated disease is a disease caused by a SARS coronavirus (SARS-CoV).

11. Use according to claim 10, wherein said SARS virus is a positive-sense single stranded RNA virus (SARS coronavirus) comprising one or more of the sequences of figure 2.

12. Use according to claim 10, wherein said SARS virus is a positive-sense single stranded RNA virus (SARS coronavirus) corresponding to GenBank accession no. AY274119 or AY278741 or AY338175 or AY338174 or AY322199 or AY 322198 or AY322197 or AH013000 or AY322208 or AY322207 or AY 322206 or AY322205 or AH012999.

13. A method for the treatment or prevention of a coronavirus associated disease in an animal, preferably a vertebrate, more preferably a bird or mammal, especially human, ape or rodent, infected with a coronavirus, said method comprising administrating interferon, to said animal, preferably a vertebrate, more preferably a bird or mammal, especially human, ape or rodent, along with a pharmaceutically acceptable carrier.

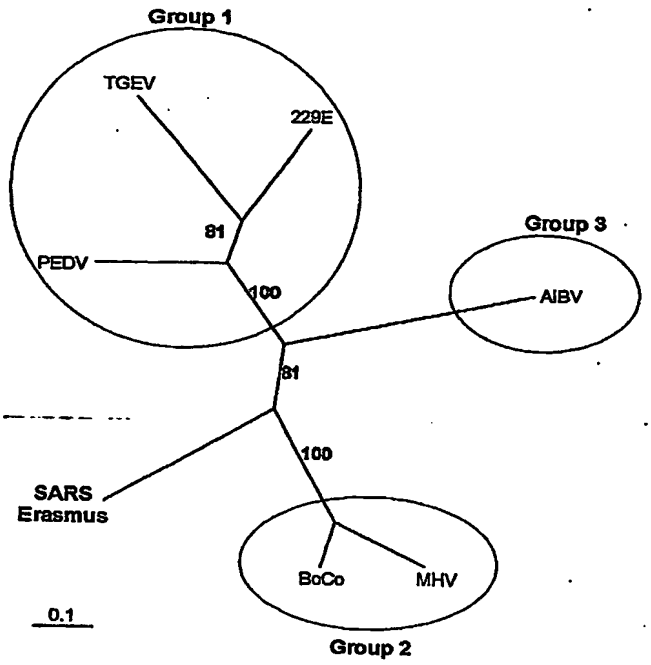
14. A method according to any of claims 13 wherein said interferon is administered together with a vaccine and/or antiviral agent.

15. A method according to claim 18, wherein said anti-viral agent is selected from the group consisting of attenuated vaccines, sub-unit vaccines, recombinant vaccines, antibody vaccines, nucleoside analogs such as ribavirin.

Title: SARS –interferon

Abstract: The invention provides use of interferon for the preparation of a medicament for the treatment or prevention of a coronavirus associated disease. Preferably said interferon is pegylated. In one aspect the invention provides a method for the treatment or prevention of a coronavirus associated disease in an animal, preferably a vertebrate, more preferably a bird or mammal, especially human, ape or rodent, infected with a coronavirus, said method comprising administrating interferon to said animal, preferably a vertebrate, more preferably a bird or mammal, especially human, ape or rodent, along with a pharmaceutically acceptable carrier. In another aspect the invention provides a method wherein said interferon is administered together with a vaccine and/or antiviral agent. Preferably said anti-viral agent is selected from the group consisting of attenuated vaccines, sub-unit vaccines, recombinant vaccines, anti-body vaccines, and nucleoside analogs such as ribovirin.

Figure 1.



5

Figure 2 RNA sequences, implied polypeptides and alignment with one close relative

EMC-1

UUGUAACUGGUGGUCUUGUACAACAGACUUCUCAGUGGUUGUCUAAUCUUUUGGGCACUACUGGUUGAAAAAC
UCAGGCCUAUCUUUGAAUGGAUUGAGGCGAAACUUAGUGCAGGAGUUGAAUUUCUCAAGGAUGCUUGGGAGAU
UCUCAAAUUUCUCAUUACAGGUGUUUUUGACAUCGUCAAGGGUCAAAUACAGGUUGCUUCAGAUAACAUCAAG
GAUUGUGUAAAAUGCUUCAUUGAUGUUGUUAACAAGGCACUCGAAAUGUGCAUUGAUCAAGUCACUAUCGCUG
GCGCAAAGUUGCGAUCACUCAACUUAGGUGAAGUCUUCAUCGCUCAAAGCAAGGGACUUUACCGUCAGUGUAU
ACGUGGCAAGGAGCAGCUGCAACUACUCAUGCCUCUUAAGGCACCAAAAGAAGUAACCUUUCU
UGAAGGUGAUUCACAUGACACAGUACUUACCUCUGAGGAGGUUGUUCUCAAGAACGGUGAA
CUCGAAGCACUCGAGACGCCCGUUGAUAGCUUCACAAAUGGAGCUAUCGUUGGCACACCAG
UCUGUGUAAAUGGCCUCAUGCUCUUAGAGAUUAAGGACAAAGAACAAUACUGCGCAUUGUC
UCCUGGUUUACUGGCUACaAACAAUGUCUUUCGCUUAAAAGGGGGUGCACCAAUUAAAGGU
GUAACCUUUGGAGAAGAUACUGUUUGGGAAGUUCAGGGUUACAAGAAUGUGAGAAUCACAU
UUGAGCUUGAUGAACGUGUUGACAAAGUGCUUAAUGAAAAGUGCUCUGUCUACACUGUUGA
AUCCGGUACCGAAGUUACUGAGUUUGCAUGUGUUGUAGCAGAGGCUGUUGUGAAGACUUUA
CAACCAGUUUCUGAUC

*Translation*Nucleotides 7 to 870: Frame 1; 288 aa

LVVLYNRLLSGCLIFWALLVEKLRPIFEWIEAKLSAGVEFLKDAWEILKFLITGVFDIVKGQIQVASDNIKDCVKCFIDVV
NKALEMCIDQVTIAGAKLRSLNLGEVFIAQSKGLYRQCIRGKEQLQLLMPLKAPKEVTFLEGDSHDTVLTSEEVVLKNGEL
EALETPVDSFTNGAIVGTPVCVNGLMLLEIKDKEQYCALSPGLLATNNVFRLKGGAPIKGVTFGEDTVWEVQGYKNVRITF
ELDERVDKVLNEKCSVYTVESGTEVTEFACVVAEAVVKTLQPVSD

Alignment

RNA-directed RNA polymerase (orf1a) murine hepatitis virus
Identities = 72/285 (25%), Positives = 118/285 (41%)

```
Query: 49   FWALLVEKLRPIFEWIEAKLSAGVEFLKDAWEILKFLITGVFDIVKGQIQVASDNIKDCV 228
            F AL V  +R I EW +  L+  +       W   + L+ G+F +  G I    + + + V
Sbjct: 638  FKALGVAVVRKITEWFD--LAVDIAASAAGWLCYQ-LVNGLFAVANGVITFVQE-VPELV 693

Query: 229  KCFIDVVNKALEMCIDQVTIA---GAKLRSLNLGEVFIAQSKGLYRQCIRGKEQLQLLMP 399
            K F+D      ++ ID ++++   G  +        V +A SK +Y   +  K      +MP
Sbjct: 694  KNFVDKFKAFFKVLIDSMSVSILSGLTVVKTASNRVCLAGSK-VYE--VVQKSLSAYVMP 750

Query: 400  LKAPKEVTFLEGDSHDTVLTSEEVVLKNGEL--EALETPVDSFTNGAIVGTPVCVNGLML 573
            +    E T L G+    V   + V +    L  +    P  SF    IV           L
Sbjct: 751  VGC-SEATCLVGEIEPAVFEDDVVDVVKAPLTYQGCCKPPTSFEKICIVDK--------L 801

Query: 574  LEIKDKEQYCAL-----SPGLLATNNVFRLKGGAPIKGVTFGEDT-VWEVQGYKNVRITF 735
               K  +Q+  +     + G+L     F   G    K V F +   V ++   + ++ITF
Sbjct: 802  YMAKCGDQFYPVVVDNDTVGVLDQCWRFPCAG----KKVEFNDKPKVRKIPSTRKIKITF 857

Query: 736  ELDERVDKVLNEKCSVYTVESGTEVTEFACVVAEAVVKTLQPVSD 870
              LD   D VL++ CS + V+    + E   VV +AV  TL P  +
Sbjct: 858  ALDATFDSVLSKACSEFEVDKDVTLDELLDVVLDAVESTLSPCKE 902
```

EMC-14

CAUCCAGCUUCUUAAGGCAGCAUAUGAAAAUUUCAAUUCACAGGACAUCUUACUUGCACCAUUGUUGUCAGCA
GGCAUAUUUGGUGCUAAACCACUUCAGUCUUUACAAGUGUGCGUGCAGACGGUUCGUACACAGGUUUAUAUUG
CAGUCAAUGACAAAGCUCUUUAUGAGCAGGUUGUCAUGGAUUAUCUUGAUAACCUGAAGCCUAGAGUGGAAGC
ACCUAAACAAGAGGAGCCACCAAACACAGAAGAUUCCAAAACUGAGGAGAAAUCUGUCGUACAGAAGCCUGUC
GAUGUGAAGCCAAAAAUUAAGGCCUGCAUUGAUGAGGUUACCACAACACUGGAAGAAACUAAGUUUCUUACCA
AUAAGUUACUCUUGUUUGCUGAUAUCAAUGGUAAGCUUUACCAUGAUUCUCAGAACAUGCUUAGAGGUGAAGA
UAUGUCUUUCCUUGAGAAGGAUGCACCUUACAUGGUAGGUGAUGUUAUCACUAGUGGUGAUAUCACUUGUGUU

Fig.2 (Cont.)

GUAAUACCCUCCAAAAAGGCUGGUGGCACUACUGAGAUGCUCUCAAGAGCUUUGAAGAAAGUGCCAGUUGAUG
AGUAUAUAACCACGUACCCUGGACAAGGAUGUGCUGGUUAUACACUUGAGGAAGCUAAGACUGCUCUUAAGAA
AUGCAAAUCUGCAUUUUAUGUACUACCUUCAGAAGCACCUAAUGCUAAGGAAGAGAUUCUAGGAACUGUAUCC
UGGAAUUGAG

Translation

Nucleotides 5 to 739: Frame 2; 245 aa

IQLLKAAYENFNSQDILLAPLLSAGIFGAKPLQSLQVCVQTVRTQVYIAVNDKALYEQVVMDYLDNLKPRVEAPKQEEPPN
TEDSKTEEKSVVQKPVDVKPKIKACIDEVTTTLEETKFLTNKLLLFADINGKLYHDSQNMLRGEDMSFLEKDAPYMVGDVI
TSGDITCVVIPSKKAGGTTEMLSRALKKVPVDEYITTYPGQGCAGYTLEEAKTALKKCKSAFYVLPSEAPNAKEEILGTVS
WN

Alignment

replicase polyprotein lab Human coronavirus 229E

Identities = 48/202 (23%); Positives = 83/202 (41%), Gaps = 13/202 (6%)
Frame = +2

```
Query: 8    LLKAAYENFNSQDILLAPLLSAGIFGAKPLQSLQVCVQTVRT---QVYIAVNDKALYEQV 178
            L+KA     N Q   L P+LS GIFG K   SL+V +    T   +V++  + +    +
Sbjct: 1371 LIKAYNTINNEQGTPLTPILSCGIFGIKLETSLEVLLDVCNTKEVKVFVYTDTEVCKVKD 1430

Query: 179  VMDYLDNLKPRVEAPKQEEPPNTEDSKTEEKSVVQKPVDVKPKIKACIDEVTTTLEETKF 358
             +  L N++ +VE PK E  P          V  KP  V  K     +++    ++
Sbjct: 1431 FVSGLVNVQ-KVEQPKIEPKP-------VSVIKVAPKPYRVDGKFSYFTEDLLCVADDKPI 1483

Query: 359  L--TNKLLLFADINGKLYHDSQNMLRG--EDMSFLEKDAP-------YMVGDVITSGDITC 508
            +  T+ +L    D     L +    +L    +D   + K  P       + +G V+    +
Sbjct: 1484 VLFTDSMLTLDDRGLALDNALSGVLSAAIKDCVDINKAIPSGNLIKFDIGSVV----VYM 1539

Query: 509  VVIPSKKAGGTTEMLSRALKKV 574
             V+PS+K        + R  +K+
Sbjct: 1540 CVVPSEKDKHLDNNVQRCTRKL 1561
```

EMC-2

UCGAGAUUUcAUcUUGACGGUGCAGGUUCUUUCACUUGACAAACUAAAGAGUCUCUUAUCCCUGCGGGAGGUU
AAGACUAUAAAAGUGUUCACAACUGUGGACAACACUAAUCUCCACACACAGCUUGUGGAUAUGUCUAUGACAU
AUGGACAGCAGUUUGGUCCAACAUACUUGGAUGGUGCUGAUGUUACAAAAAUUAAACCUCAUGUAAAUCAUGA
GGGUAAGACUUUCUUUGUACUACCUAGUGAUGACACACUACGUAGUGAAGCUUUCGAGUACUACCAUACUCUU
GAUGAGAGUUUUCUUGGUAGGUACAUGUCUGCUUUAAACCACACAAAGAAAUGGAAA

Translation

Nucleotide 2 to 349: Frame 2; 116 aa

RDFILTVQVLSLDKLKSLLSLREVKTIKVFTTVDNTNLHTQLVDMSMTYGQQFGPTYLDGADVTKIKPHVNHEGKTFFVLP
SDDTLRSEAFEYYHTLDESFLGRYMSALNHTKKWK

Alignment

> Bovine Coronavirus RNA-Dependent RNA polymerase

Identities = 25/90 (27%), Positives = 44/90 (48%)
Frame = +2

```
Query: 80   IKVFTTVDNTNLHTQLVDMSMTYGQQFGPTYLDGADVTKIKPHVNHEGKTFFVLPSDDTL 259
            + +  TVD  N   + V +  ++G+  G  + DG +VTK K  +N++GK FF   +  +
Sbjct: 1565 VDILLTVDGVNFTNRFVPVGESFGKSLGNVFCDGVNVTKHKCDINYKGKVFFQFDNLSSE 1624

Query: 260  RSEAFEYYHTLDESFLGRYMSALNHTKKWK 349
              +A       D+  L  Y + L +  KW+
Sbjct: 1625 DLKAVRSSFNFDQKELLAYYNMLVNCSKWQ 1654
```

Fig 2. (Cont.)

EMC13:

CUGAAGAAGUAGUGGaAAAUCCUACCAUACAGAAGGAAGUCAUAGAGUGUGACGUGAAAACUACCGAAGUUGU
AGGCAAUGUCAUACUUAAACCAUCAGAUGAAGGUGUUAAAGUAACACAAGAGUUAGGUCAUGAGGAUCUUAUG
GCUGCUUAUGUGGAAAACACAAGCAUUACCAUUAAGAAACCUAAUGAGCUUUCACUAGCCUUAGGUUUAAAAA
CAAUUGCCACUCAUGGUAUUGCUGCAAUUAAUAGUGUUCCUUGGAGUAAAAUUUUGGCUUAUGUCAAACCAUU
CUUAGGACAAGCAGCAAUUACAACAUCAAAUUGCGCUAAGAGAUUAGCACAACGUGUGUUUAACAAUUAUAUG
CCUUAUGUGUUUACAUUAUUGUUCCAAUUGUGUACUUUUACUAAAAGUACCAAUUCUAGAAUUAGAGCUUCAC
UACCUACAACUAUUGCUAAAAAUAGUGUUAAGAGUGUUGCUAAAUUAUGUUUGGAUGCCGGCAUUAAUUAUGU
GAAGUCACCCAAAUUUUCUAAAUUGUUCACAAUCGCUAUGUGGCUAUUGUUGUUAAGUAUUUGCUUAGGUUCU
CUAAUCUGUGUAACUGCUGCUUUUGGUGUACUCUUAUCUAAUUUUGGUGCUCCUUCUUAUUGUAAUGGCGUUA
GAGAAUUGUAUCUUAAUUCGUCUAACGUUACUACUAUGGAUUUCUGUGAAGGUUCUUUUCCUUGCAGCAUUUG
UUUAAGUGGAUUAGACUCCCUUGAUUCUUAUCCAGCUCUUGAAACCAUUCAGGUGACGAUUUCAUCGUACAAG
CUAGACUUGACAAUUUUAGGUCUGGCCGCUG

Translation

>~out: 3 to 833: Frame 3 277 aa
EEVVENPTIQKEVIECDVKTTEVVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALGLKTIATHGIAA
INSVPWSKILAYVKPFLGQAAITTSNCAKRLAQRVFNNYMPYVFTLLFQLCTFTKSTNSRIRASLPTTIAKNSVKSVAKLC
LDAGINYVKSPKFSKLFTIAMWLLLLSICLGSLICVTAAFGVLLSNFGAPSYCNGVRELYLNSSNVTTMDFCEGSFPCSIC
LSGLDSLDSYPALETIQVTISSYKLDLTILGLAA

Alignment

bovine coronavirus RNA-dependent RNA Polymerase
Identities = 50/269 (18%),

```
Query: 57   KTTEVVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALGLKTIATH- 233
            K  +V  +VI+       +K  + L   D+   ++       ++  N+LS+A+ + TI
Sbjct: 2046 KPFKVEDSVIVNDDTSEIKYVKSLSIVDVYDMWLTGCRYVVRTANDLSMAVNVPTIRKFI 2105

Query: 234  --GIAAINSVPWSKI-LAYVKPFLGQAAITTSNCAKRLAQRVFN--NYMPYVFTLLF--- 389
              G+  + S+P   + L  +KP          N  K +  ++    N++ ++F LLF
Sbjct: 2106 KFGMTLV-SIPIDLLNLREIKPVF--------NVVKAVRNKISACFNFIKWLFVLLFGWI 2156

Query: 390  -----QLCTFTKSTNSRIRASLPTTIAKNSVKSVAKLCLDAGINYVKSPKFSKLFTIAMW 554
                     +T    S++   L     KN+  +     +  G   + +           +W
Sbjct: 2157 KISADNKVIYTTEVASKLTCKLVALAFKNAFLTFKWSVVARGACIIAT-------IFLLW 2209

Query: 555  XXXXXXXXXXXXXVTAAFGVLLSNFGAPSYCNGVRELYLNSSNVTTM------------- 695
                        G L      P++   + +   ++ ++ T+
Sbjct: 2210 FNFIYANVIFSDFYLPKIGFL------PTFVGKIAQWIKSTFSLVTICDLYSIQDVGFKN 2263

Query: 696  DFCEGSFPCSICLSGLDSLDSYPALETIQ 782
              +C GS  C  CL+G D LD+Y A++ +Q
Sbjct: 2264 QYCNGSIACQFCLAGFDMLDNYKAIDVVQ 2292
```

EMC-3

GUGGUAAGAUUGUUAGUACUUGUUUUAAACUUAUGCUUAAGGCCACAUUAUUGUGCGUUCU
UGCUGCAUUAGUUUGUUAUAUCGUUAUGCCAGUACAUACAUUGUCAAUCCAUGAUGGUUAC
ACAAAUGAAAUCAUUGGUUACAAAGCCAUUCAGGAUGGUGUCACUCGUGACAUCAUUUCUA
CUGAUGAUUGUUUUGCAAAUAAACAUGCUGGUUUUGACGCAUGGUUUAGCCAGCGUGGUGG
UUCAUACAAAAAUGACAAAAGCUGCCCUGUAGUAGCUGCUAUCAUUACAAGAGAGAUUGGU
UUCAUAGUGCCUGGCUUACCGGGUACUGUGCUGAGAGCAAUCAAUGGUGACUUCUUGCAUU
UCCUACCUCGUGUUUUUAGUGCUGUUGGCAACAUUUGCUACACACCUUCCAAACUCAUUGA
GUAUAGUGAUUUUGCUACCUCU

Translation

Nucleotide 3-449; 149 aa
GKIVSTCFKLMLKATLLCVLAALVCYIVMPVHTLSIHDGYTNEIIGYKAIQDGVTRDIISTDDCFANKHAGFD
AWFSQRGGSYKNDKSCPVVAAIITREIGFIVPGLPGTVLRAINGDFLHFLPRVFSAVGNICYTPSKLIEYSDF
ATS

Fig 2 (Cont.)

Alignment

> Murine Hepatitis Virus RNA-Dependent RNA polymerase

```
Identities = 48/126 (38%),

Query: 78   YIVMPVHTLSIHDGYTNEIIGYKAIQDGVTRDIISTDDCFANKHAGFDAWFSQRGG--SY 251
            + +MP + +   D         +K I +GV RD+  TD CFANK    FD W+    G   Y
Sbjct: 2859 WALMPTYAVHKSDMQLPLYASFKVIDNGVLRDVSVTDACFANKFNQFDQWYESTFGLAYY 2918

Query: 252  KNDKSCPVVAAIITREIGFIVPGLPGTVLRAINGDFLHFLPRVFSAVGNICYTPSKLIEY 431
            +N K+CPVV A+I ++IG  +  +P TVLR      LHF+   F+     CYTP   I Y
Sbjct: 2919 RNSKACPVVVAVIDQDIGHTLFNVPTTVLR-YGFHVLHFITHAFATDSVQCYTPHMQIPY 2977

Query: 432  SDFATS 449
             +F  S
Sbjct: 2978 DNFYAS 2983
```

EMC-4

ACAGACAUCAAUCACUUCUGCUGUUCUGCAGAGUGGUUUUAGGAAAAUGGCAUUCCCGUCAGGCAAAGUUGAA
GGGUGCAUGGUACAAGUAACCUGUGGAACUACAACUCUUAAUGGAUUGUGGUUGGAUGACACAGUAUACUGUC
CAAGACAUGUCAUUUGCACAGCAGAAGACAUGCUUAAUCCUAACUAUGAAGAUCUGCUCAUUCGCAAAUCCAA
CCAUAGCUUUCUUGUUCAGGCUGGCAAUGUUCAACUUCGUGUUAUUGGCCAUUCUAUGCAAAAUUGUCUGCUU
AGGCUUAAAGUUGAUACUUCUAACCCUAAGACACCCAAGUAUAAAUUUGUCCGUAUCCAACCUGGUCAAACAU
UUUCAGUUCUAGCAUGCUACAAUGGUUCACCAUCUGGUGUUUAUCAGUGUGCCAUGAGACCUAAUCAUACCAU
UAAAGGUUCUUUCCUUAAUGGAUCAUGUGGUAGUGUUGGUUUUAACAUUGAUUAUGAUUGCGUGUCUUUCUGC
UAUAUGCAUCAUAUGGAGCUUCCAACAGGAGUACACGCUGGUACUGACUUAGAAGGUAAAUUCUAUGGUCCAU
UUGUUGACAGACAAACUGCACAGGCUGCAGGUACAGACACAACCAUAACAUUAAAUGUUUUGGCAUGGCUGUA
UGCUGCUGUUAUCAAUGGUGAUA

Translation

Nucleotides 2 to 679: Frame 2; 226 aa

```
QTSITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNPNYEDLLIRKSNHSFLVQAG
NVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKFVRIQPGQTFSVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNI
DYDCVSFCYMHHMELPTGVHAGTDLEGKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVINGD
```

Alignment

RNA-directed RNA polymerase murine hepatitis virus

```
Identities = 122/222 (54%)

Query: 8    SITSAVLQSGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDTVYCPRHVICTAEDMLNP 187
            S+T++ LQSG  KM  P+ KVE C+V VT G  TLNGLWLDD VYCPRHVIC++ DM +P
Sbjct: 3326 SVTTSFLQSGIVKMVSPTSKVEPCIVSVTYGNMTLNGLWLDDKVYCPRHVICSSADMTDP 3385

Query: 188  NYEDLLIRKSNHSFLVQAGNVQLRVIGHSMQNCLLRLKVDTSNPKTPKYKFVRIQPGQTF 367
            +Y +LL R ++  F V +G + L V+ + MQ C L L V   NP TPKY F  ++PG+TF
Sbjct: 3386 DYPNLLCRVTSSDFCVMSGRMSLTVMSYQMQGCQLVLTVTLQNPNTPKYSFGVVKPGETF 3445

Query: 368  SVLACYNGSPSGVYQCAMRPNHTIKGSFLNGSCGSVGFNIDYDCVSFCYMHHMELPTGVH 547
            +VLA YNG P G +    +R +HTIKGSFL GSCGSVG+ +  D V F YMH +EL TG H
Sbjct: 3446 TVLAAYNGRPQGAFHVTLRSSHTIKGSFLCGSCGSVGYVLTGDSVRFVYMHQLELSTGCH 3505

Query: 548  AGTDLEGKFYGPFVDRQTAQAAGTDTTITLNVLAWLYAAVIN 673
              GTD  G FYGP+ D Q  Q    D T T+NV+AWLYAA+ N
Sbjct: 3506 TGTDFSGNFYGPYRDAQVVQLPVQDYTQTVNVVAWLYAAIFN 3547
```

EMC-5

Note that this sequence is not fully in frame.

AGUUGGAAAAGAUGGCAGAUCAGGCUAUGACCCAAAUGUACAAACAGGCAAGAUCUGAGGA
CAAGAGGGCAAAAGUAACUAGUGCUAUGCAAACAAUGCUCUUCACUAUGCUUAGGAAGCUU
GAUAAUGAUGCACUUAACAACAUUAUCAACAAUGCGCGUGAUGGUUGUGUUCCACUCAACA
UCAUACCAUUGACUACAGCAGCCAAACUCAUGGUUGUUGUCCCUGAUUAUGGUACCUACAA
GAACACUUGUGAUGGUAACACCUUUACAUAUGCAUCUGCACUCUGGGAAAUCCAGCAAGUU
GUUGAUGCGGAUAGCAAGAUUGUUCAACUUAGUGAAAUUAACAUGGACAAUUCACCAAAUU
UGGCUUGGCCCCUUAUUGUUACAGCUCUAAGAGCCAACUCAGCUGUUAAACUACAGAAUAA
UGAACUGAGUCCAGUAGCACUACGACAGAUGUCCUGUGCGGCUGGUACCACACAAACAGCU
UGUACUGAUGACAAUGCACUUGCCUACUAUAACAAUUCGAAGGGAGGUAGGUUUGUGCUGG

Fig 2. (Cont.)

CAUUACUAUCAGACCACCAAGAUCUCAAAUGGGCUAGAUUCCCUAAGAGUGAUGGUACAGG
UACAAUUUACACAGAACUGGAACCACCUUGUAGGUUUGUUACAGACACACCAAAAGGGCCU
AAAGUGAAAUACUUGUACUUCAUCAAGGCUUAAACAACCUAAAUAGAGGUAUGGUGCUGGG
CAGUUUAGCUGCUACAGUACGUCUUCAGGCUGGAAAUGCUACAGAAGUaCCUGCCAAUUCA
ACUGUGCUUUCCUUCUGUGCUUUUGCAGUAGACCCUGCUAAAGCAUAUaAAGGAUUACCUA
GCAAGUGGAGGACAACCAAUCACCAACUGUGUGAAGAUGUUGUGUACACACACUGGUACAG
GACAGGCAAUUACUGUAACACCAGAAGCUAACAUGGACCAAGAGUCCUYUGGUGGUGCUUC
AUGUUGUCUGUAUUGUAGAUGCCACAUUGACCAUCCAAAUCCUAAAGGAYUCUGUGACUUG
AAAGGUAAGUACGUCCAAAUACCUACCACUUGUGCUAAUGACCCAGUGGGUUUUACACUUA
GAAACACAGUCUGUACCGUCUGCGGAAUGUGGAAAGGUUAUGGCUGUAGUUGUGACCAACU
CCGCGAACCCUUGAUGCAGUCUGCGGAUGCAUCAMCGUUUUUAAACGGGUUUGCGGUGUAA
GUGCAGCCCGUCUUACACCGUGCGGCACAGGCACUAGUACUGAUGUCGUCUACAGGGCUUU
UGAUAUUUACAACGAAAAAGUUGCUGGUUYUGCAAAGUUCCUAAAAACUAA

Translation 1
Nucleotide 3-701 ; 233 aa
LEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKLDNDALNNIINNARDGCVPLNIIPLTTAAKLMVV
VPDYGTYKNTCDGNTFTYASALWEIQQVVDADSKIVQLSEINMDNSPNLAWPLIVTALRANSAVKLQNNELSP
VALRQMSCAAGTTQTACTDDNALAYYNNSKGGRFVLALLSDHQDLKWARFPKSDGTGTIYTELEPPCRFVTDT
PKGPKVKYLYFIKA

Translation 2
FKRVCGVSA-ARLTPCGTGTSTDVVYRAFDIYNEKVAGXAKFLK

Alignment 1 of translation 1 sequence
RNA-Dependent RNA Polymerase: bovine coronavirus
Identities = 181/413 (43%),

```
Query: 3     LEKMADQAMTQMYKQARSEDKRAKVTSAMQTMLFTMLRKXXXXXXXXXXXXXXRDGCVPLN 182
              LE+MAD A+T MYK+AR  DK++KV SA+QTMLF+M+RK                GCVPLN
Sbjct: 3985  LERMADLALTNMYKEARINDKKSKVVSALQTMLFSMVRKLDNQALNSILDNAVKGCVPLN 4044

Query: 183   IIPLTTAAKLMVVVPDYGTYKNTCDGNTFTYASALWEIQQVVDADSKIVQLSEINMDNSP 362
               IP   A  L ++VPD   Y   D    TYA  +W+IQ + D+D    QL+EI+ D +
Sbjct: 4045  AIPSLAANTLTIIVPDKSVYDQVVDNVYVTYAGNVWQIQTIQDSDGTNKQLNEISDDCN- 4103

Query: 363   NLAWPLIVTALRAN--SAVKLQNNELSPVALRQMSCAAGTTQTACTDDNALAYYNNSKGG 536
                 WPL++ A R N  SA  LQNNEL P  L+     +G  QT  T      YYNNS  G
Sbjct: 4104  ---WPLVIIANRHNEVSATVLQNNELMPAKLKTQVVNSGPDQTCNTPTQ--CYYNNSNNG 4158

Query: 537   RFVLALLSDHQDLKWARFPKSDGTGTIYTELEPPCRFVTDTPKGPKVKYLYFIKA*TT*I 716
              + V A+LSD   LK+ +  K DG   +  EL+PPC+F     KG K+KYLYF+K   T
Sbjct: 4159  KIVYAILSDVDGLKYTKILKDDG-NFVVLELDPPCKFTVQDVKGLKIKYLYFVKGCNTLA 4217

Query: 717   EVWCWAV*LLQYVFRL------EMLQKYLPIQLCFPSVLLQ*TLLKHIKDYLASGGQPIT 878
                 W V  +  , RL      E       + LC  SV  + T L    D++  GG PI
Sbjct: 4218  R--GWVVGTISSTVRLQAGTATEYASNSSILSLCAFSVDPKKTYL----DFIQQGGTPIA 4271

Query: 879   NCVKMLCTHTGTGQAITVTPEANMDQESXGGASCCLYCRCHIDHPNPKGXCDLKGKYVQI 1058
              NCVKMLC H GTG AITV P+A  +Q+S GGAS C+YCR  ++HP+  G C L+GK+VQ+
Sbjct: 4272  NCVKMLCDHAGTGMAITVKPDATTNQDSYGGASVCIYCRARVEHPDVDGLCKLRGKFVQV 4331

Query: 1059  PTTCANDPVGFTLRNTVCTVCGMWKGYGCSCDQLREPLMQSADASXFLNGFAV 1217
              P     DPV + L + VC VCG W+   CSC    +  +QS D + FLNGF V
Sbjct: 4332  PVG-IKDPVSYVLTHDVCQVCGFWRDGSCSCVS-TDTTVQSKDTN-FLNGFGV 4381
```

Alignment 2 of translation 2 sequence
RNA-directed RNA polymerase (ORF1B) [murine hepatitis virus]

Identities = 24/44 (54%),

```
Query: 1199  FKRVCGVSA-ARLTPCGTGTSTDVVYRAFDIYNEKVAGXAKFLK 1327
              FKRV G S  ARL PC +G  TDV  RAFDI N   AG   + K
Sbjct: 18    FKRVRGTSVNARLVPCASGLDTDVQLRAFDICNANRAGIGLYYK 61
```

Fig 2 (Cont.)

Note that this sequence is not fully in frame.

UGACAUCUUACGCGUAUAUGCUAACUUAGGUGAGCGUGUACGCCAAUCAUUAUUAAAGACU
GUACAAUUCUGCGAUGCUAUGCGUGAUGCAGGCAUUGUAGGCGUACUGACAUUAGAUAAUC
AGGAUCUUAAUGGGAACUGGUACGAUUUCGGUGAUUUCGUACAAGUAGCACCAGGCUGCGG
AGUUCCUAUUGUGGAUUCAUAUUACUCAUUGCUGAUGCCCAUCCUCACUUUGaCUAGGgCA
UUGGCUGCUGAGUCCcAUAUGGAUGCUGAUCUCGCAAAaCCACUUAUUAaGUGGgAUUUGC
UGAAACAUGAUUUUACGGAAGAGAGACUUUGUCUCUUCGACCGUUAUUUUAAAUAUUGGGA
CCAGACAUACCAUCCCAAUUGUAUUAACUGUUUGGAUGAUAGGUGUAUCCUUCAUUGUGCA
AaCUUUAAUGUGUUAUUUUCUACUGUGUUUcCACCUACAAGUUUUGGACCACUAGUAAGAA
AAAUAUUUGUAGAUGGUGUUCCUUCUGUUGUUUCAACUGGAUACCAUUUUCGUGAGUUAGG
AGUCGUACAUAAUCAGGAUGUAAACUUACAUAGCUCGCGUCUCAGUUUCAAGGAACUUUUA
GUGUAUGCUGCUGAUCCAGCUAUGCAUGCAGCUUCUGGCAAUUAAUUGCUAGAUAAACGCA
CUACAUGCUUUUCAGUAGCUCCACUAACAAACAAUGUUGCUUUUCAAACUGUCAAACCCGG
UAAUUUUAAUAAAGACUUUUAUGACUUUGCUGUGUCUAAA

Translation 1

Nucleotide 2 to 652: Frame 2; 217 aa

DILRVYANLGERVRQSLLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGCGVPIVDSYYSLLM
PILTLTRALAAESHMDADLAKPLIKWDLLKHDFTEERLCLFDRYFKYWDQTYHPNCINCLDDRCILHCANFNV
LFSTVFPPTSFGPLVRKIFVDGVPSVVSTGYHFRELGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGN

Translation 2

656 to 772: Frame 2; 39 aa

LLDKRTTCFSVAPLTNNVAFQTVKPGNFNKDFYDFAVSK

Alignment

ORF1ab polyprotein Murine hepatitis virus

Identities = 157/257 (61%),

```
Query: 2    DILRVYANLGERVRQSLLKTVQFCDAMRDAGIVGVLTLDNQDLNGNWYDFGDFVQVAPGC 181
            DI+ VY  LG     ++LL T +F DA+ +AG+VGVLTLDNQDL G WYDFGDFV+  PGC
Sbjct: 4626 DIINVYKKLGPIFNRALLNTAKFADALVEAGLVGVLTLDNQDLYGQWYDFGDFVKTVPGC 4685

Query: 182  GVPIVDSYYSLLMPILTLTRALAAESHMDADLAKPLIKWDLLKHDFTEERLCLFDRYFKY 361
            GV + DSYYS +MP+LT+  AL +E  ++     +    +DL+++DFT+ +L LF +YFK+
Sbjct: 4686 GVAVADSYYSYMMPMLTMCHALDSELFVNGTYRE----FDLVQYDFTDFKLELFTKYFKH 4741

Query: 362  WDQTYHPNCINCLDDRCILHCANFNVLFSTVFPPTSFGPLVRKIFVDGVPSVVSTGYHFR 541
            W  TYHPN   C DDRCI+HCANFN+LFS V P T FGPLVR+IFVDGVP VVS GYH++
Sbjct: 4742 WSMTYHPNTCECEDDRCIIHCANFNILFSMVLPKTCFGPLVRQIFVDGVPFVVSIGYHYK 4801

Query: 542  ELGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGN*LLDKRTTCFSVAPLTNNVAFQT 721
            ELGVV N DV+ H  RLS K+LL+YAADPA+H AS + LLD RT CFSVA +T+ V FQT
Sbjct: 4802 ELGVVMNMDVDTHRYRLSLKDLLLYAADPALHVASASALLDLRTCCFSVAAITSGVKFQT 4861

Query: 722  VKPGNFNKDFYDFAVSK 772
            VKPGNFN+DFY+F +SK
Sbjct: 4862 VKPGNFNQDFYEFILSK 4878
```

EMC-7 .

ACCUUCAGAAUUAUGGUGAAAAUGCUGUUAUACCAMAAGGAAUAAUGAUGAAUGUCGCAAAGUAUACUCAACU
GUGUCAAUACUUAAAUACACUUACUUUAGCUGUACCCUACAACAUGAGAGUUAUUCACUUUGGUGCUGGCUCU
GAUAAAGGAGUUGCACCAGGUACAGCUGUGCUCAGACAAUGGUUGCCAACUGGCACACUACUUGUCGAUUCAG
AUCUUAAUGACUUCGUCUCCGACGCAGAUUCUACUUUAAUUGGAGACUGUGCAACAGUACAUACGGCUAAUAA
AUGGGACCUUAUUAUUAGCGAUAUGUAUGACCCUAGGACCAAACAUGUGACAAAAGAGAAUGACUCUAAAGAA
GGGUUUUUCACUUAUCUGUGUGGAUUUAUAAAGCAAAAACUAGCCCUGGGUGGUUCUAUAGCUGUAAAGAUAA
CAGAGCAUUCUUGGAAUGCUGACCUUUACAAGCUUAUGGGCCAUUUCUCAUGGUGGACAGCUUUUGUUACAAA
UGUAAAUGCAUCAUCAUCGGAAGCAUUUUUAAUUGGGGCUAACUAUCUUGGCAAGCCGAAGGAACAAAUUGAU
GGCUAUACCAUGCAUGCUAACUACAUUUUCUGGAGGAACACAAAUCCUAUCCAGUUGUCUUCCUAUUCACUCU
UUGACAUGAGCAAAUUUCCUCUUAAAUUAAGAGGAACUGCUGUAAUGUCUCUUAAGGAGAAUCAAAUCAAUGA
UAUGAUUUAUUCUCUUCUGGAAAAAGGUAGGCUUAUCAUUAGAGAAAACAACAGAGUUGUGGUUUCAAGUGAU
AUUCUUGUUAACAACUAAACGAACAUGUUUAUUUUCUUAUUAUUUCUUACUCUCACUAGUGGUAGUGACCUUG
ACCGGUGCACCACUUUUGAUGAUGUUCAAGCUCCUAAUUACACUCAACAUACUUCAUCUAUGAGGGGGGUUUA
CUAUCCUGAUGAAAUUUUUAGAUCAGACACUCUUUAUUUAACUCAGGAUUUAUUUCUUCCAUUUUAUUCUAAU
GUUACAGGGUUUCAUACUAUUAAUCAUACGUUUGGCAACCCUGUCAUACCUUUUAAGGAUGGUAUUUAUUUUG
CUGCCACAGAGAAAUCAAAUGUUGUCCGUGGUUGGGUUUUUGGUUCUACCAUGAACAACAAGUCACAGUCGGU

Fig 2. (Cont.)

GAUUAUUAUUAACAAUUCUACUAAUGUUGUUAUACGAGCAUGUAACUUUGAAUUGUGUGACAACCCUUUCUUU
GCUGUUUCUAAACCCAUGGGUACACAGACACAUACUAUGAUAUUCGAUAAUGCAUUUAAUUGCACUUUCGAGU
ACAUAUCUGAUGCCUUUUCGCUUGAUGUUUCAGAAAAGUCAGGUAAUUUUAAACACUUACGAGAGUUUGUGUU
UAAAAAUAAAGAUGGGUUUCUCUAUGUUUAUAAGGGCUAUCAACCUAUAGAUGUAGUUCGUGAUCUACCUUCU
GGUUUUAACACUUUGAAACCUAUUUUUAAGUUGCCUCUUGGUAUUAACAUUACAAAUUUUAGAGCCAUUCUUA
CAGCCUUUUCACCUGCUCAAGACAUUUGGGGCACGUCAGCUGCAGCCUAUUUUGUUGGCUAUUUAAAGCCAAC
UACAUUUAUGCUCAAGUAUGAUGAAAAUGGUACAAUCACAGAUGCUGUUGAUUGUUCUCAAAAUCCACUUGCU
GAACUCAAAUGCUCUGUUAAGAGCUUUGAGAUUGACAAAGGAAUUUACCAGACCUCUAAUUUCAGGGUUGUUC
CCUCAGGAGAUGUUGUGAGAUUCCCUAAUAUUACAAACUUGUGUCCUUUUGGAGAGGUUUUUAAUGCUACUAA
AUUCCCUUCUGUCUAUGCAUGGGAGAGAAAAAAAAUUUCUAAUUGUGUUGCUGAUUACUCUGUGCUCUACAAC
UCAACAUUUUUUUCAACCUUUAAGUGCUAUGGCGUUUCUGCCACUAAGUUGAAUGAUCUUUGCUUCUCCAAUG
UCUAUGCAGAUUCUUUUGUAGUCAAGGGAGAUGAUGUAAGACAAAUAGCGCCAGGACAAACUGGUGUUAUUGC
UGAUUAUAAUUAUAAAUUGCCAGAUGAUUUCAUGGGUUGUGUCCUUGCUUGGAAUACUAGGAACAUUGAUGCU
ACUUCAACUGGUAAUUAUAAUUAUAAAUAUAGGUAUCUUAGACAUGGCAAGCUUAGGCCCUUUGAGAGAGACA
UAUCUAAUGUGCCUUUCUCCCCCUGAUGGCAAACCUUGCACCCCACCUGCUCUUAAUUGUUAUUGGCCAUUAAA
UGAUUAUGGUUUUUACACCACUACUGGCAUUGGCUACCAACCUUACAGAGUUGUAGUACUUUCUUUUGAACUU
UUAAAUGCACCGGCCACGGUUUGUGGACCAAAAUUAUCCACUGACCUUAUUAAGAACCAGUGUGUCAAUUUUA
AUUUUAAUGGACUCACUGGUACUGGUGUGUUAACUCCUUCUUCAAAGAGAUUUCAACCAUUUCAACAAUUUGG
CCGUGAUGUUUCUGAUUUCACUGAUUCCGUUCGAGAUCCUAAAACAUCUGAAAUAUUAGACAUUUCACCUUGC
UCUUUUGGGGGUGUAAGUGUAAUUACACCUGGAACAAAUGCUUCAUCUGAAGUUGCUGUUCUAUAUCAAGAUG
UUAACUGCACUGAUGUUUCUACAGCAAUUCAUGCAGAUCAACUCACACCAGCUUGGCGCAUAUAUUCUACUGG
AAACAAUGUAUUCCAGACUCAAGCAGGCUGUCUUAUAGGAGCUGAGCAUGUCGACACUUCUUAUGAGUGCGAC
AUUCCUAUUGGAGCUGGCAUUUGUGCUAGUUACCAUACAGUUUCUUUAUUACGUAGUACUAGCCAAAAAUCUA
UUGUGGCUUAUACUAUGUCUUUAGGUGCUGAUAGUUCAAUUGCUUACUCUAAUAACACCAUUGCUAUACCUAC
UAACUUUUCAAUUAGCAUUACUACAGAAGUAAUGCCUGUUUCUAUGGCUAAAACCUCCGUAGAUUGUAAUAUG
UACAUCUGCGGAGAUUCUACUGAAUGUGCUAAUUUGCUUCUCCAAUAUGGUAGCUUUUGCACACAACUAAAUC
GUGCACUCUCGUGGUAUUGCUGCUGAACAGGAUCGCAACACAC

Translation 1

Nucleotides 3 to 818: Frame 3 272 aa (orf 1ab)

LQNYGENAVIPQGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLRQWLPTGTLLVDSDLNDFVSDA
DSTLIGDCATVHTANKWDLIISDMYDPRTKHVTKENDSKEGFFTYLCGFIKQKLALGGSIAVKITEHSWNADLYKLMGHFS
WWTAFVTNVNASSSEAFLIGANYLGKPKEQIDGYTMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKENQIND
MIYSLLEKGRLIIRENNRVVVSSDILVNN

Translation 2

Nucleotide 828 to 3089: Frame 3 756 aa (S protein)

MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFLPFYSNVTGFHTINHTFGNPVI
PFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINNSTNVVIRACNFELCDNPFFAVSKPMGTQTHTMIFDNAFNCTFE
YISDAFSLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQD
IWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNPLAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVRFPNITNLCP
FGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVI
ADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTT
TGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTPSSKRFQPFQQFGRDVSDFTDSVRDPKT
SEILDISPCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYEC
DIPIGAGICASYHTVSLLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISITTEVMPVSMAKTSVDCNMYICGDST
ECANLLLQYGSFCTQLNRALSWYCC

Alignment 1 of translation 1

replicase [bovine coronavirus]

Identities = 183/271 (67%),

```
Query: 3     LQNYGENAVIPQGIMMNVAKYTQLCQYLNTLTLAVPYNMRVIHFGAGSDKGVAPGTAVLR 182
              L NYG+    +P G MMNVAKYTQLCQYLNT TLAVP NMRV+H GAGS+KGVAPG+AVLR
Sbjct: 6822  LWNYGKPVTLPTGCMMNVAKYTQLCQYLNTTTLAVPVNMRVLHLGAGSEKGVAPGSAVLR 6881

Query: 183   QWLPTGTLLVDSDLNDFVSDADSTLIGDCATVHTANKWDLIISDMYDPRTKHVTKENDSK 362
              QWLP GT+LVD+DL  FVSD+ +T  GDC T+     +WDLIISDMYDP TK++ + N SK
Sbjct: 6882  QWLPAGTILVDNDLYPFVSDSVATYFGDCITLPFDCQWDLIISDMYDPITKNIGEYNVSK 6941

Query: 363   EGFFTYLCGFIKQKLALGGSIAVKITEHSWNADLYKLMGHFSWWTAFVTNVNASSSEAFL 542
              +GFFTY+C  I+ KLALGGS+A+KITE SWNA+LYKLMG+F++WT F TN NASSSE FL
```

Fig 2. (Cont.)

```
Sbjct: 6942 DGFFTYICHMIRDKLALGGSVAIKITEFSWNAELYKLMGYFAFWTVFCTNANASSSEGFL 7001

Query: 543  IGANYLGKPKEQIDGYTMHANYIFWRNTNPIQLSSYSLFDMSKFPLKLRGTAVMSLKENQ 722
            IG NYLGKPK +IDG  MHANY+FWRN+       +YSLFDM+KFPLKL GTAV++L+ +Q
Sbjct: 7002 IGINYLGKPKVEIDGNVMHANYLFWRNSTVWNGGAYSLFDMAKFPLKLAGTAVINLRADQ 7061

Query: 723  INDMIYSLLEKGRLIIRENNRVVVSSDILVN 815
            INDM+YSLLEKG+L++R+ N+ V   D LVN
Sbjct: 7062 INDMVYSLLEKGKLLVRDTNKEVFVGDSLVN 7092
```

Alignment 2 (Spike protein of coronavirus)
E2 glycoprotein precursor - murine hepatitis virus (strain JHM); contains spike glycoprotein

Identities = 199/798 (24%), Positives = 314/798 (39%), Gaps = 48/798 (6%)
Frame = +3

```
Query: 828  MFIFLLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSM-------------RGVYYP-DEI 965
            +F+F+L L    G   D    F  +Q  NY  + +S                RG YY  D +
Sbjct: 2    LFVFILLLPSCLGYIGD----FRCIQTVNYNGNNASAPSISTEAVDVSKGRGTYYVLDRV 57

Query: 966  FRSDTLYLTQDLFLPF----YSNV--TGFHTINHTFGNP--VIPFKDGIYFAATE-KSNV 1118
            + + TL LT   + P     Y N+  TG +T++ T+   P  +  F DGI+       K+N
Sbjct: 58   YLNATLLLTG--YYPVDGSNYRNLALTGTNTLSLTWFKPPFLSEFNDGIFAKVQNLKTNT 115

Query: 1119 VRGW--------VFGSTMNNKXXXXXXXXXXXXXXXRACNFELCDNPFFAVSKPMGTQTHT 1277
              G         V GS   N                      C + +C  P+    KP
Sbjct: 116  PTGATSYFPTIVIGSLFGNTSYTVVLEPYNNIIMASVCTYTICQLPY-TPCKP-------- 167

Query: 1278 MIFDNAFNCTFEYISDAFSLDVSEKSGNFKHLREFVFKNKDGFLYVY---KGYQPIDVVR 1448
                  N     +   +  DV        K  R F F      +LY +    +G
Sbjct: 168  --------NTNGNRVIGFWHTDVKPPICLLK--RNFTFNVNAPWLYFHFYQQGGTFYAYYA 218

Query: 1449 DLPSGFNTLKPIFKLPLGINITNFRAILTAFSPAQDIWGTSAAAYFVGYLKPTTFMLKYD 1628
            D PS    L   F + +G  +T +  +     +P       T A  Y+V  L    ++  ++
Sbjct: 219  DKPSATTFL---FSVYIGDILTQYFVLPFICTPTAG--STLAPLYWVTPLLKRQYLFNFN 273

Query: 1629 ENGTITDAVDCSQNPLAELKCSVKSFEIDKGIYQTSNFRVVPSGDVVR-FPNITNLCPFG 1805
            E G IT AVDC+ + ++E+KC  +S     G+Y  S + V P G V R  PN+ + C
Sbjct: 274  EKGVITSAVDCASSYISEIKCKTQSLLPSTGVYDLSGYTVQPVGVVYRRVPNLPD-CKIE 332

Query: 1806 EVFNATKFPSVYAWERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADS 1985
            E   A   PS   WER+   NC  + S L       +  C  + A+K+  +CF +V  D
Sbjct: 333  EWLTAKSVPSPLNWERRTFQNCNFNLSSLLRYVQAESLSCNNIDASKVYGMCFGSVSVDK 392

Query: 1986 FVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHG 2165
            F +       +  G +G +   NYK+      C L ++    + T   NYN       R+G
Sbjct: 393  FAIPRSRQIDLQIGNSGFLQTANYKIDTAATSCQLYYSLPKNNVT-INNYNPSSWNRRYG 451

Query: 2166 KLRPFERDISNVPFSPDGKPCTPPALNCYWPLNDYGFYTTTGIGYQPYRVVVLSFELLNA 2345
                                        + +ND                R  + +  LLN
Sbjct: 452  ------------------------------FKVND---------------RCQIFANILLNG 468

Query: 2346 --PATVCGPKL---STDLIKNQCVNFNFNGLTGTGVLTP-SSKRFQPFQQFGRDVSDFTD 2507
               T C   L   +T++    CV ++  G+TG GV      +  +  +Q    DV+   +
Sbjct: 469  INSGTTCSTDLQLPNTEVATGVCVRYDLYGITGQGVFKEVKADYYNSWQALLYDVNGNLN 528

Query: 2508 SVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLYQDVNCTDVSTAIHADQLTPAWR 2687
              RD  T++   I  C  G VS        + E A+LY+++NC+ V T   +  +  P
Sbjct: 529  GFRDLTTNKTYTIRSCYSGRVSAAY--HKEAPEPALLYRNINCSYVFTNNISREENPL-- 584

Query: 2688 IYSTGNNVFQTQAGCLIGAEH--VDTSYECDIPIGAGICASYHTVSLLR---STSQK--S 2846
              N F +  GC++ A++    +    C++ +GAG+C  Y      R   ST  +  +
Sbjct: 585  ------NYFDSYLGCVVNADNRTDEALPNCNLRMGAGLCVDYSKSRRARRSVSTGYRLTT 638

Query: 2847 IVAYTMSLGADSSIAYSN-NTIAIPTNFSISITTEVMPVSMAKTSVDCNMYICGDSTECA 3023
              Y   L  DS  +      + IPTNF+I    E + +   K ++DC  ++CGD+  C
Sbjct: 639  FEPYMPMLVNDSVQSVGGLYEMQIPTNFTIGHHEEFIQIRAPKVTIDCAAFVCGDNAACR 698

Query: 3024 NLLLQYGSFCTQLNRALS 3077
```

```
                L++YGSFC  +N  L+
Sbjct: 699  QQLVEYGSFCDNVNAILN 716
```

RDG1 seq

```
5
UUCAAAGCcUUCAAACNUAUGUAACACAACAACUAAUCAGGGMUGcUGAAAUCHCGSCUUCUGCUAAUCUUGC
UGCUACUAAAAUGUCUGAGUGUGUUCUUGGACAAUCAAAAAGAGUUGACUUUUGUGGAAAGGGCUACCACCUU
AUGUCCUUCCCACAAGCAGCCCCGCAUGGUGUUGUCUUCCUACAUGUCACGUAUGUGCCAUCCCAGGAGAGGA
ACUUCACCACAGCGCCAGCAAUUUGUCAUGAAGGCAAAGCAUACUUCCCUCGUGAAGGUGUUUUUGUGUUUAA
10 UGGCACUUCUUGGUUUAUUACACAGAGGAACUUCUUUUCUCCACAAAUAAUUACUACAGACAAUACAUUUGUC
UCAGGAAAUUGUGAUGUCGUUAUUGGCAUCAUUAACAACACAGUUUAUGAUCCUCUGCAACCUGAGCUUGACU
CAUUCAAAGAAGAGCUGGACAAGUACUUCAAAAAUCAUACAUCACCAGAUGUUGAUCUUGGCGACAUUUCAGG
CAUUAACGCUUCUGUCGUCAACAUUCAAAAAGAAAUUGACCGCCUCAAUGAGGUCGCUAAAAAUUUAAAUGAA
UCACUCAUUGACCUUCAAGAAUUGGGAAAAUAUGAGCAAUAUAUUAAGUGgCCCUGGUACGUCUGGGU
15
```

Translation 1
Nucleotides 3 to 650: Frame 3; 216 aa

```
QSLQXYVTQQLIRXAEIXXSANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAIC
HEGKAYFPREGVFVFNGTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPD
20 VDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVW
```

Translation 2
Nucleotides 37 to 339: Frame 1; 101 aa

```
SGXLKXXLLLILLLLKCLSVFLDNQKELTFVERATTLCPSHKQPRMVLSSYMSRMCHPRRGTSPQRQQFVMKAKHTSLVKV
25 FLCLMALLGLLHRGTSFLHK
```

Translation 3
Nucleotides 343 to 576: Frame 1; 78 aa

```
LLQTIHLSQEIVMSLLASLTTQFMILCNLSLTHSKKSWTSTSKIIHHQMLILATFQALTLLSSTFKKKLTASMRSLKI
30
```

Alignment of translation 1

```
S glycoprotein [murine hepatitis virus]
          Length = 1376

35  Identities = 86/218 (39%), Positives = 129/218 (59%), Gaps = 3/218 (1%)
    Frame = +3

Query: 6    SLQTYVTQQLIRXAEIXXSANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVF 185
             +L  Y+++QL     I  SA  A  K++ECV  Q+ R++FCG G H++S  Q AP+G+ F
40  Sbjct: 1105 ALNAYISKQLSDSTLIKFSAAQAIEKVNECVKSQTTRINFCGNGNHILSLVQNAPYGLYF 1164

Query: 186  LHVTYVPSQERNFTTAPAICHEG-KAYFPREGVFVFNGTSWFITQRNFFSPQIITTDNTF 362
             +H +YVP+       +P +C  G +   P+ G FV +   W  T  +++ P+ IT  N+
Sbjct: 1165 IHFSYVPTSFTTANVSPGLCISGDRGLAPKAGYFVQDDGEWKFTGSSYYYPEPITDKNSV 1224
45
Query: 363  VSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTS--PDVDLGDISGINASVVNI 536
             V  +C V        + +   P L  FKEELDK+FKN TS  PD+ L D   +N + +++
Sbjct: 1225 VMSSCSVNYTKAPEVLLNSSIPNLPDFKEELDKWFKNQTSIAPDLSL-DFEKLNVTFLDL 1283

50  Query: 537  QKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVW 650
               E++R+ E  K LNES I+L+E+G YE Y+KWPWYVW
    Sbjct: 1284 SDEMNRIQEAIKKLNESYINLKEVGTYEMYVKWPWYVW 1321
```

55 **EMC-8**

```
AGGCCAAAACAGCGCCGACCCCAAGGUUUACCCAAUAAUACUGCGUCUUGGUUCACAGCUCUCACUCAGCAUG
GCAAGGAGGAACUUAGAUUCCCUCGAGGCCAGGGCGUUCCAAUCAACACCAAUAGUGGUCCAGAUGACCAAAU
UGGCUACUACCGAAGAGCUACCCGACGAGUUCGUGGUGGUGACGGCAAAAUGAAAGAGCUCAGCCCCAGAUGG
UACUUCUAUUACCUAGGAACUGGCCCAGAAGCUUCACUUCCCUACGGCGCUAACAAAGAAGGCAUCGUAUGGG
60 UUGCAACUGAGGGAGCCUUGAAUACACCCAAAGACCACAUUGGCACCCGCAAUCCUAAUAACAAUGUUGCC

*Translation*
Nucleotides 1 to 363: Frame 1; 121 aa
RPKQRRPQGLPNNTASWFTALTQHGKEELRFPRGQGVPINTNSGPDDQIGYYRRATRRVRGGDGKMKELSPRWYFYYLGTG
65 PEASLPYGANKEGIVWVATEGALNTPKDHIGTRNPNNNXA
```

Fig 2. (Cont.)

Alignment
nucleocapsid protein - bovine coronavirus (strain Mebus)

Identities = 55/129 (42%),

```
Query: 1   RPKQRRPQGLPNNTA-----SWFTALTQHGK-EELRFPRGQGVPINTNSGPDDQIGYYRR 162
           +PKQ     LP+         SWF+ +TQ  K +E  F  GQGVPI       +  GY+ R
Sbjct: 44  QPKQTATSQLPSGGNVVPYYSWFSGITQFQKGKEFEFAEGQGVPIAPGVPATEAKGYWYR 103

Query: 163 ATRR-VRGGDGKMKELSPRWYFYYLGTGPEASLPYGANKEGIVWVATEGA-LNTPKDHIG 336
             RR  +  DG  ++L PRWYFYYLGTGP A   YG + +G+ WVA+  A +NTP D I
Sbjct: 104 HNRRSFKTADGNQRQLLPRWYFYYLGTGPHAKDQYGTDIDGVFWVASNQADVNTPAD-IL 162

Query: 337 TRNPNNNXA 363
           R+P+++ A
Sbjct: 163 DRDPSSDEA 171
```

EMC-11: unknown sequence

```
UUGCAUACCGCAAUGUUCUUCUUCGUAAGAACGGUaAUAAGGGAGCCGGUGGUCAUAGCUgUGGCAUGAUCUA
AAGUCUUAUGACUUAGGUGACGAGCUUGGCACUGAUCCCAUUGAAGAUUAUGAACAAAACUGGAACACUAAGC
AUGGCAGUGGUGCACUCCGUGAACUCACUCGUGAGCUCAAUGGAGGUGCAGUCACUCGCUAUGUCGACAACAA
UUUCUGUGGCCCAGAUGGGUACCCUCUUGAUUGCAUCAAAGAUUUUCUCGCACGCGCGGGCAAGUCAAUGUGC
ACUCUUUCCGAACAACUUGAUUACAUCGAGUCGaAGAGAGGUGUCUACUGCUGCCGUGACCAUGAGCAUGAAA
UUGCCUgGGUUCACUGAGCGCUCUGAUAAGAGCUACGAGCACCAGACACCCUUCGaAAUUAAGAGUGCCAAGA
AAaUUGACACUUUCAAAAGGGGAAUGCCCCAAAGCUUGUGUUUCCUCUUAACUCAAAAGUCAAAGUCAUUCAA
CCACGUGUUGAAAAGAAAAAGACUGAGGGUUUCAUGGGGCGUAUACGCUCUGUGUACCCUGUUGCAUCUCCAC
AGGAGUGUAACAAUAUGCACUUGUCUACCUUGAUGAAAUGUAAUCAUUGCGAUGAAGCUUCAUGGCAGACGUG
CGACUUUCUGAAAGCCACUUGUGAACAUUGUGGCACUGAAAAUUUAGUUAUUGAAGGACCUAGUACAUGUGGG
UACCUACCUACUAAUGCUGUAGUGAAAAUGCCAUGUCCUGCCUGUCAAGACCCAGAGAUUGGACCUGAGCAUA
GUGUUGCAGAUUAUCACAACCACUCAAACAUUGAAACUCGACUCCGCAAGGGAGGUAGGACUAGAUGUUUUGG
AGGCUGUGUGUUUGCCUAUGUUGGCUGCUAUAAUAAGCGUGCCUACUGGGUUCCUCGUGCUAGUGCUGAUAUU
GGCUCAGGCCAUACUGGCAUUACUGGUGACAAUGUGGAGACCUUGAAUGAGGAUCUCCUUGAGAUACUGAGUC
GUGAACGUGUUAACAUUAACAUUGUUGGCGAUUUUCAUUUGAAUGAAGAGGUUGCCAUCAYUUUGGCAUCYUU
CUCUGCUUCUACAAGUGCCUUUAUUGACACUAUAAAGAGUCUUGAUUACAAGUCUUUCAAAACCAUUGUUGAG
UCCUGCGGUAACUAUAAAGUUACCAAGGGAAAGCCCGUAAAAGGUGCUUGGAACAUUGGACAACAGAGAUCAG
UUUUAACACCACUGUGUGGUUUUCCCUCACAGGCUGCUGGUGUUAUCAGAUCAAUUUUUGCGCGCACACUUGA
UGCAGCAAACCACUCAAUUCCUGAUUUGCAAAGAGCAGCUGUCACCAUACUUGAUGGUAUUUCUGAACAGUCA
UUACGUCUUGUCGACGCCAUGGUUUAUACUUCAGACCUGCUCACCAACAGUGUCAUUAUUAUGGCAUAUGUAA
CUGGUGGUCUUGUACAACAGACU
```

Translation of putative open reading frames

```
>~out: 78 to 1: Frame -2        26 aa
DFRSCHSYDHRLPYYRSYEEEHCGMQ
>~out: 59 to 379: Frame 2       107 aa
LWHDLKSYDLGDELGTDPIEDYEQNWNTKHGSGALRELTRELNGGAVTRYVDNNFCGPDGYPLDCIKDFLARAGKSMCTLS
EQLDYIESKRGVYCCRDHEHEIAWVH
>~out: 283 to 89: Frame -1       65 aa
LARACEKIFDAIKRVPIWATEIVVDIASDCTSIELTSEFTECTTAMLSVPVLFIIFNGISAKLVT
>~out: 90 to 614: Frame 3       175 aa
VTSLALIPLKIMNKTGTLSMAVVHSVNSLVSSMEVQSLAMSTTISVAQMGTLLIASKIFSHARASQCALFPNNLITSSRRE
VSTAAVTMSMKLPGFTERSDKSYEHQTPFEIKSAKKIDTFKRGMPQSLCFLLTQKSKSFNHVLKRKRLRVSWGVYALCTLL
HLHRSVTICTCLP
>~out: 204 to 124: Frame -2       27 aa
RVTAPPLSSRVSSRSAPLPCLVFQFCS
>~out: 312 to 208: Frame -2       35 aa
SSCSERVHIDLPARARKSLMQSRGYPSGPQKLLST
>~out: 485 to 258: Frame -3       76 aa
EETQALGHSPFESVNFLGTLNFEGCLVLVALIRALSEPRQFHAHGHGSSRHLSSTRCNQVVRKECTLTCPRVRENL
>~out: 397 to 287: Frame -1       37 aa
LLSERSVNPGNFMLMVTAAVDTSLRLDVIKLFGKSAH
>~out: 364 to 486: Frame 1        41 aa
NCLGSLSALIRATSTRHPSKLRVPRKLTLSKGECPKACVSS
>~out: 490 to 401: Frame -1       30 aa
VKRKHKLWGIPLLKVSIFLALLISKGVWCS
>~out: 446 to 1483: Frame 2       346 aa
```

HFQKGNAPKLVFPLNSKVKVIQPRVEKKKTEGFMGRIRSVYPVASPQECNNMHLSTLMKCNHCDEASWQTCDFLKATCEHC
GTENLVIEGPSTCGYLPTNAVVKMPCPACQDPEIGPEHSVADYHNHSNIETRLRKGGRTRCFGGCVFAYVGCYNKRAYWVP
RASADIGSGHTGITGDNVETLNEDLLEILSRERVNINIVGDFHLNEEVAIXLAXFSASTSAFIDTIKSLDYKSFKTIVESC
GNYKVTKGKPVKGAWNIGQQRSVLTPLCGFPSQAAGVIRSIFARTLDAANHSIPDLQRAAVTILDGISEQSLRLVDAMVYT
SDLLTNSVIIMAYVTGGLVQQT
>~out: 643 to 494: Frame -1 50 aa
SFIAMITFHQGRQVHIVTLLWRCNRVHRAYTPHETLSLFLFNTWLNDFDF
>~out: 627 to 511: Frame -2 39 aa
LHFIKVDKCILLHSCGDATGYTERIRPMKPSVFFFSTRG
>~out: 704 to 612: Frame -3 31 aa
LNFQCHNVHKWLSESRTSAMKLHRNDYISSR
>~out: 774 to 631: Frame -2 48 aa
QAGHGIFTTALVGRYPHVLGPSITKFSVPQCSQVAFRKSHVCHEASSQ
>~out: 826 to 737: Frame -1 30 aa
VVVIICNTMLRSNLWVLTGRTWHFHYSISR
>~out: 863 to 744: Frame -3 40 aa
SYLPCGVEFQCLSGCDNLQHYAQVQSLGLDRQDMAFSLQH
>~out: 756 to 992: Frame 3 79 aa
KCHVLPVKTQRLDLSIVLQIITTTQTLKLDSAREVGLDVLEAVCLPMLAAIISVPTGFLVLVLILAQAILALLVTMWRP
>~out: 952 to 830: Frame -1 41 aa
ANISTSTRNPVGTLIIAANIGKHTASKTSSPTSLAESSFNV
>~out: 1056 to 922: Frame -2 45 aa
KSPTMLMLTRSRLSISRRSSFKVSTLSPVMPVWPEPISALARGTQ
>~out: 1237 to 956: Frame -1 94 aa
SLLSNVPSTFYGLSLGNFIVTAGLNNGFERLVIKTLYSVNKGTCRSREXCQXDGNLFIQMKIANNVNVNTFTTQYLKEILI
QGLHIVTSNASMA
>~out: 1140 to 1060: Frame -2 27 aa
SRLFIVSIKALVEAEXDAKXMATSSFK
>~out: 1131 to 1205: Frame 3 25 aa
RVLITSLSKPLLSPAVTIKLPRESP
>~out: 1410 to 1183: Frame -2 76 aa
TMASTRRNDCSEIPSSMVTAALCKSGIEWFAASSVRAKIDLITPAACEGKPHSGVKTDLCCPMFQAPFTGFPLVTL
>~out: 1186 to 1311: Frame 1 42 aa
SYQGKARKRCLEHWTTEISFNTTVWFSLTGCWCYQINFCAHT
>~out: 1283 to 1191: Frame -3 31 aa
HQQPVRENHTVVLKLISVVQCSKHLLRAFPW
>~out: 1248 to 1457: Frame 3 70 aa
HHCVVFPHRLLVLSDQFLRAHLMQQTTQFLICKEQLSPYLMVFLNSHYVLSTPWFILQTCSPTVSLLWHM
>~out: 1381 to 1482: Frame 1 34 aa
TVITSCRRHGLYFRPAHQQCHYYGICNWWSCTTD

EMC12: unknown sequence

UGCUUGCUCAUGCUGAAGAGACAAGAAAAUUAAUGCCUAUAUGCAUGGAUGUUAGAGCCAU
AAUGGCAACCAUCCAACGUAAGUAUAAAGGAAUUAAAAUUCAAGAGGGCAUCGUUGACUAU
GGUGUCCGAUUCUUCUUUUAUACUAGUAAAGAGCCUGUAGCUUCUAUUAUUACGAAGCUGA
ACUCUCUAAAUGAGCCGCUUGUCACAAUGCCAAUUGGUUAUGUGACACAUGGUUUUAAUCU
UGAAGAGGCUGCGCGCUGUAUGCGUUCUCUUAAAGCUCCUGCCGUAGUGUCAGUAUCAUCA
CCAGAUGCUGUUACUACAUAUAAUGGAUACCUCACUUCGUCAUCAAAGACAUCUGAGGAGC
ACUUUGUAGAAACAGUUUCUUUGGCUGGCUCUUACAGAGAUUGGUCCUAUUCAGGACAGCG
UACAGAGUUAGGUGUUGAA

Translation of putative open reading frames

>~out: 3 to 446: Frame 3 148 aa
LAHAEETRKLMPICMDVRAIMATIQRKYKGIKIQEGIVDYGVRFFFYTSKEPVASIITKLNSLNEPLVTMPIGYVTHGFNL
EEAARCMRSLKAPAVVSVSSPDAVTTYNGYLTSSSKTSEEHFVETVSLAGSYRDWSYSGQRTELGVE
>~out: 100 to 11: Frame -2 30 aa
ILIPLYLRWMVAIMALTSMHIGINFLVSSA
>~out: 188 to 33: Frame -1 52 aa
RVQLRNNRSYRLFTSIKEESDTIVNDALLNFNSFILTLDGCHYGSNIHAYRH
>~out: 64 to 159: Frame 1 32 aa
WQPSNVSIKELKFKRASLTMVSDSSFILVKSL
>~out: 220 to 143: Frame -2 26 aa
PIGIVTSGSFREFSFVIIEATGSLLV
>~out: 293 to 192: Frame -1 34 aa
HYGRSFKRTHTARSLFKIKTMCHITNWHCDKRLI
>~out: 397 to 224: Frame -2 58 aa
EPAKETVSTKCSSDVFDDEVRYPLYVVTASGDDTDTTAGALRERIQRAASSRLKPCVT
>~out: 229 to 288: Frame 1 20 aa

Fig 2. (Cont.)

HMVLILKRLRAVCVLLKLLP
>~out: 292 to 372: Frame 1 27 aa
CQYHHQMLLLHIMDTSLRHQRHLRSTL
>~out: 444 to 340: Frame -3 35 aa
QHLTLYAVLNRTNLCKSQPKKLFLQSAPQMSLMTK
>~out: 416 to 351: Frame -1 22 aa
IGPISVRASQRNCFYKVLLRCL
>~out: 365 to 445: Frame 2 27 aa
GALCRNSFFGWLLQRLVLFRTAYRVRC
>~out: 376 to 435: Frame 1 20 aa
KQFLWLALTEIGPIQDSVQS

Figure 3.

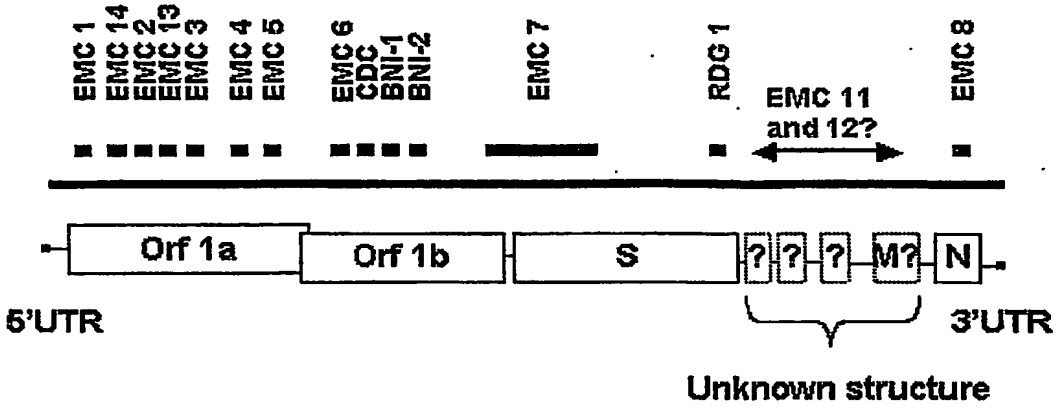

EMC 1
EMC 14
EMC 2
EMC 13
EMC 3
EMC 4
EMC 5
EMC 6
CDC
BNI-1
BNI-2
EMC 7
RDG 1
EMC 11
and 12?
EMC 8
Orf 1a
Orf 1b
S
?
?
?
M?
N
5'UTR
3'UTR
Unknown structure

Figure 4.

Comparison of N-termini of the S proteins of the group 2 coronaviruses

HCV OC43	MFLILLISLPTAFAVIG**DL-KCTT**VSINDID
MHV A59	MLFVFILFLPSCLGYIG**DF-RC**IQLVNSNGA
BCV	MFLILLISLPMAFAVIG**DL-KCTT**VSINDVD
SARS	MF-IFLLFL-TLTSG-SDLDRCTTFDDVQAP

Figure 5.

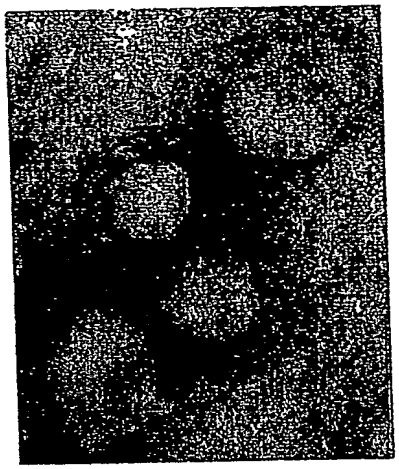


Figure 6.